Functional Genomics of Ankylosing Spondylitis

Anna Sanniti
Merton College
University of Oxford

A thesis submitted in partial
fulfilment of the requirements for the degree of
Doctor of Philosophy
Trinity Term, 2017

Abstract

This thesis aimed to determine the functional genomic changes specific to Ankylosing Spondylitis (AS) for different disease relevant immune cell types and relate this to disease associated genetic variants. AS is a highly heritable inflammatory disease mainly affecting the spine. Functional genomic approaches at the transcriptomic and epigenetic level in the context of disease are accelerating the understanding of the functional consequences of genetic variation, and how this may impact disease susceptibility and aetiology. Disease-associated genetic variants may be active only in certain cell types; therefore understanding the underlying functional mechanisms requires investigation in relevant tissues and cells, and in the disease state. TNF-α inhibition as a current biologic therapy has proved very effective in the majority but not all AS patients. To understand this response further, cell type specific changes in treatment responders were also investigated.

The transcriptomic profile in 15 AS patients with active disease is reported here for six immune blood cell types ($CD4^+$ and $CD8^+$ T cells, $CD14^+$ monocytes, $CD16^+$ neutrophils, NK cells and $CD19^+$ B cells) using RNA-seq, in purified cell types for messenger RNAs, long non-coding RNAs and microRNAs. Enrichment for NF-κB signalling and HDAC Class I signalling was found across cell types, most significantly in $CD14^+$ monocytes and $CD16^+$ neutrophils. The gene expression profile 3 months after commencing anti-TNF therapy is reported for three cell types, providing insights into the molecular basis of treatment response. *NEAT1*, a long non-coding RNA was found to be correlated with disease activity following anti-TNF therapy (P=0.031) in $CD14^+$ monocytes.

Changes in chromatin accessibility are informative for putative regulatory DNA elements and may inform hypotheses for molecular mechanisms modulated by genetic variation. Here, the chromatin accessibility landscapes in 5 AS patients are established for three cell types using an Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq). In $CD14^+$ monocytes, 437 genes were found differentially expressed in AS patients in both RNA-seq and ATAC-seq (observed enrichment P=2.2x10^{-16}). Following TNF-α inhibition therapy, differentially expressed genes in both datasets were found in $CD8^+$ T cells (P= 0.02). *PTGER4*, a GWAS reported gene was found down-regulated in $CD8^+$ T cells from treated AS patients (FDR=0.0002). Integration of ATAC-seq data with fine-mapped SNPs also revealed a differentially closed chromatin peak in one patient at rs4672505 (locus 2p15), specifically in $CD8^+$ T cells. This SNP may explain the association signal found at this locus, with evidence of regulating *B3GNT2*. Future work aims to validate these findings in a larger cohort and collect AS specific epigenetic marks using ChIP-seq, along with functional validation of rs4672505 using Capture-C and CRISPR.

Acknowledgements

First and foremost I would like to express my gratitude for the opportunity to work on this project. I am thankful for the support and encouragement provided by my supervisors Professor Julian Knight and Professor Paul Bowness, along with funding provided by Arthritis Research UK.

A special appreciation goes to my colleague and friend Alicia Lledó Lara, with whom this thesis would not be possible. I would like to express my gratitude for her continuous support throughout the project, both academically and emotionally through long days (and nights!) of sample processing. I would also like to extend my gratitude to all other members of the Knight Group, especially to Dr Hai Fang, Dr Andrew Brown, Giuseppe Scozzafava, Katie Burnham, Dr Gabriele Migliorini and former lab members Dr Benjamin Fairfax, Sara Danielli, Dr Emma Davenport, Evelyn Chan and Dr Katharine Doherty. Another special mention goes to Dr Adrián Cortés, whose patience and continued willingness to help has been invaluable. This thesis would also not be possible without the research nurses Karen Doig and Jennifer O'Donoghue, who were always extremely organised and so helpful with patient recruitment.

Declarations

I declare that unless otherwise stated, all work presented in this thesis is my own. The GENExpress ID Study was established in 2013, in collaboration with the Botnar Research Centre and Nuffield Orthopedic Centre.

Part of the patient recruitment and sample collection was carried out by former lab members Dr Katharine Doherty and Evelyn Chan. The samples collected and processed by myself were part of a collaborative effort with current lab members Alicia Lledó Lara, Giuseppe Scozzafava and Dr Andrew Brown. RNA extraction for a batch of healthy controls was carried out by Dr Andrew Brown. RNA-seq and ATAC-seq data was generated by the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics. Advice for data analysis was provided by Dr Helen Lockstone, Alicia Lledó Lara, Dr Adrián Cortés and Dr Gabriele Migliorini.

Data, resources and advice for the fine-mapping analysis was provided by Dr Adrián Cortés. The script for the TSS plot was provided by Dr Silvia Salatino from the Oxford Genomics Centre. Dr Hai Fang provided the function to create KEGG pathways with gene expression data.

Contents

List of Figures

List of Tables

Abbreviations

AS	Ankylosing Spondylitis
ASAS	Assessment of SpondyloArthritis international Society
ASDAS	ASAS-endorsed Disease Activity Score
ATAC-seq	Assay For Transposase-Accessible Chromatin sequencing
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
CD	Crohn's Disease
CDCV	Common Disease Common Variant
CDVR	Common Disease Rare Variant
ChIP-seq	Chromatin Immunoprecipitation and DNA sequencing
CRISPR	Clustered Regularly-Interspaced Short Palindromic Repeats
CTCF	CCCTC-binding Factor
CNV	Copy Number Variant
DHS	DNase I Hyper-Sensitivity
DMARDs	Disease Modifying Anti-Rheumatic Drugs
DNase-seq	DNase I hypersensitive sites sequencing
dsQTL	dnase I sensitivity quantitive trait locus
eQTL	expression Quantitative Trait Loci
ENCODE	Encyclopaedia of DNA Elements
EWAS	Epigenome Wide Association Studies
FAIRE-seq	Formaldehyde-Assisted Isolation of Regulatory Elements followed by Sequencing
FDR	False Discovery Rate
GO	Gene Ontology
GWAS	GWAS Genome-Wide Association Studies
GTeX	Genotype Tissue Expression
GSEA	Gene Set Enrichment Analysis
HAT	Histone AcetylTransferase
HDAC	Histone DeACetylase
HLA-B27	Human Leukocyte Antigen B27
IBD	Inflammatory Bowel Disease
IDR	Irreproducible Discovery Rate
INDELs	Insertion and Deletions
INF-γ	Interferon-gamma
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KIR	Killer Immunoglobulin-like Receptor
lncRNA	long non-coding RNA
HMM	Hidden Markov Model

IO	Immune Ontology
LCR	Locus Control Region
LD	Linkage Disequilibrium
LPS	LipoPolySaccharide
lncRNA	long non-coding RNA
MAF	Minor Allele Frequency
MAX	MYC Associated Factor X
meQTL	methylation Quantitative Trait Loci
MHC	Major Histocompatibility Complex
miRNA	micro RNA
mRNA	messenger RNA
NGS	Next-Generation Sequencing
NK	Natural Killer
NSAIDs	Non Steroidal Anti-Inflammatory Drugs
OR	Odds Ratio
PBCs	Peripheral Blood cells
PBMCs	Peripheral Blood Mononuclear Cell
PC	Principle Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PICS	Probabilistic Identification of Causal SNPs
PMNs	Polymorphonuclear Leukocytes
PP	Posterior Probability
PsA	Psoriatic Arthritis
QC	Quality Control
RA	Rheumatoid Arthritis
ReA	Reactive Arthritis
RNA-seq	RNA Sequencing
ZMIZ1	Zinc Finger MIZ-Type Containing 1
YY1	Yin And Yang 1 Protein
SLE	Systemic Lupus Erythematosus
SNP	Single Nucleotide Polymorphism
SpA	Spondyloarthritis
TAD	Topologically Associated Domain
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TNF	Tumour Necrosis Factor-alpha
TSS	Transcription Start Site
UC	Ulcerative Colitis
UTR	UnTranslated Region
WTCCC	Wellcome Trust Case Control Consortium

Chapter 1

Introduction

1.1 Aims and objectives

The overall aim of this thesis is to understand the cell specific changes in gene expression and chromatin accessibility occurring in ankylosing spondylitis (AS). Genome-wide association studies have provided us with a large number of genetic variants that highlight specific loci implicated in AS pathogenesis. In combination with the functional genomic approaches adopted in this thesis, these variants may begin to inform us about the genes and pathways relevant to the pathogenesis of AS and the molecular mechanisms by which potential functional variants may act. In this chapter, an overview of current understanding of AS pathogenesis is presented, together with the genetic basis of disease and functional genomic approaches to assaying gene expression and gene regulation.

1.2 Ankylosing Spondylitis

AS is a heritable chronic inflammatory condition affecting the spine and sacroiliac joints, causing eventual bone apposition, ankylosis (fusion) and loss of mobility. It is one of a group of disorders described as Spondyloarthropathies (SpA), which also includes a subset of psoriatic arthritis (PsA), reactive arthritis (ReA) and spondylitis associated with inflammatory bowel disease (IBD). Men are three times more likely to be affected with AS than women in Europe, and the

disease commonly develops between the ages of 15 and 40 (Dean 2016; Brown et al. 2002).

The prevalence of AS is between 0.1-2%, with the highest frequency in Haida Americans. Understanding the epidemiology of AS may be important in understanding any aetiological triggers or risk factors that may be driving the disease. For example, the Human Leukocyte Antigen B27 (HLA-B27), is a Major Histocompatibility Complex (MHC) Class I molecule associated with AS, and is one of the strongest genetic associations to a common disease to date. The geographic distribution of AS and HLA-B*27 frequency is consistent with this, as populations with the highest allele frequency of HLA-B*27 have the highest disease prevalence. One exception to this rule was found among West Africans: although the percentage of HLA-B*27 positive individuals was high, the risk of developing AS was much lower than in Caucasians (Brown et al. 1997). This suggests the presence of a non-B*27 protective genetic factor, or environmental triggers that may also have a large part to play in AS pathogenesis (Brown et al. 2002; Brown et al. 1997).

1.2.1 Ankylosing Spondylitis and Spondyloarthritis

SpA refers to a spectrum of immune-mediated inflammatory diseases with overlapping features, which differ from other types of inflammatory arthritis in genetic predisposition, pathogenesis and outcome. This group of diseases is the second most prevalent form of inflammatory arthritis, with a prevalence similar to that of rheumatoid arthritis (Stolwijk et al. 2015b). Collectively, the SpAs may affect 0.6-1% of the adult US population (Reveille 2011b). This prevalence is higher in northern populations and lower in Africa (Brown et al. 1997). Many patients with SpA often exhibit multiple phenotypes over time. Approximately 7% of spondyloarthitis patients often go on to develop gut inflammation and IBD, especially Crohn's disease (CD) (Mielants et al. 1995).

Many AS patients have a similar bowel phenotype to those suffering from CD (Baeten et al. 2002; Mielants et al. 1995), suggesting that these diseases may be different manifestations of the same pathophysiological condition. SpAs can be considered as a single disease due to the sharing of genetic, immunopathological and structural features (Baeten et al. 2013). SpA subtypes such as IBD and psoriasis are often found within a single family (Costantino et al. 2017). Recent studies have demonstrated that the spondyloarthropathies have a large amount of genetic overlap, and may share mechanisms and pathways linked to cytokine production, aberrant peptide presentation and the elimination of intracellular bacteria (Costello et al. 2015; Evans et al. 2011; Kleinschek et al. 2009; Ciccia et al. 2009). This conclusion is currently incomplete due to the lack of knowledge around the further cellular and molecular mechanisms that drive SpA pathogenesis. The differences between peripheral and axial arthritis are still unknown, but it may be the case that different molecular mechanisms underlie these subtypes, since peripheral and axial arthritis respond differently to conventional therapies (Zochling et al. 2006). Classifying diseases based on phenotypic presentation may limit research into underlying pathophysiology and the use of effective treatments to specific disease subtypes. If the same cellular and molecular mechanisms are affecting different organs or tissues, it may be more beneficial for research and treatments to consider the phenotypes as belonging to a single disease (Baeten et al. 2013).

1.3 Pathophysiology of Ankylosing Spondylitis

1.3.1 Clinical presentations

AS is an arthritic condition affecting primarily the spine, with onset predominately in males and usually before the age of 40 (Reveille 2013; Brown et al. 2002). AS patients also often exhibit extra-axial manifestations, such

as uveitis, psoriasis, enthesitis and gut inflammation (Stolwijk et al. 2015b). Some clinical presentations are quite distinct from other related inflammatory diseases, with the characteristic "bamboo spine" which occurs due to the bridging syndesmophyte formation between vertebrae. This causes significant pain, posture and economic problems to the patient (Boonen and Severens 2002).

1.3.2 Clinical assessment

A major challenge in spondyloarthritis is being able to diagnose or predict the development of disease. Assessment of disease activity is used to describe the level of inflammation being experienced, and is most often assessed by the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI). This index is a self assessment consisting of a list of six questions, shown in Table 1.1, and is also the criterion used in the assessment of disease activity for the AS patients in this study. Other assessment indexes of disease activity also exist, more recently the Assessment of SpondyloArthritis international Society (ASAS)-endorsed disease activity score (ASDAS), which was demonstrated to perform better than BASDAI or single-item variables in discriminating between high and low disease activity state according to doctors as well as patients (Lukas et al. 2009). However, the ASDAS was not found to be superior to BASDAI in its ability to discriminate between high and low disease activity states in axial psoriatic arthritis (Eder et al. 2010). Disease activity is also often assessed by MRI and radiological scoring systems (Braun et al. 2002). The Bath Ankylosing Spondylitis Radiology Index (BASRI) is an index used for grading radiographic changes in patients with AS (MacKay et al. 1998). AS can also be classified into early or late onset disease (Montilla et al. 2012; Hmamouchi et al. 2011), although this is often not a criteria is the diagnosis of AS and disease severity.

BASDAI
How would you describe the overall level of:
tiredness/fatigue you have experienced?
neck, back or hip pain you have had?
pain/swelling in joints other than the neck, back or hips you have had?
discomfort you have had from any areas tender to touch or pressure?
morning stiffness you have had since the time you wake up?
How active was your spondylitis on average during the last week?

Table 1.1: The Bath Ankylosing Sponsylitis Disease Activity Index (BASDAI). A validated disease activity index: the criterion is rated on a visual analogue score of 0-10.

1.3.3 Disease aetiology

AS is characterised as an immune mediated inflammatory disease. The adaptive versus innate immune response is emerging as an important contributor to the aetiology of AS and other SpAs (Inman 2009; Masters et al. 2009), and possibly may have more of a role than adaptive responses (Inman and El-Gabalawy 2009; Lories and Baeten 2009).

Understanding exactly how HLA-B*27 is directly involved in AS pathogenesis remains an area of active research despite this genetic association having been recognised for nearly 40 years. The original observation of HLA-B*27 by Brewerton and colleagues was reported in 1973 (Brewerton et al. 1973). The specific subtype showing strongest association varies by population but includes HLA-B*2705 (Khan et al. 2007; Reveille 2006) (see section 1.5.2). The classical MHC class I molecules are expressed on almost all nucleated cells, with B*27 being one of the most common B alleles in caucasians (Bowness 2015). The main function of HLA-B*27 is antigen presentation to the T cell receptor of cytotoxic

$CD8^+$ T cells. The original hypothesis of HLA-B*27 induced SpA is that the disease is mediated by $CD8^+$ T lymphocytes responding to peptides presented by HLA-B*27 molecules (Benjamin and Parham 1992; Benjamin and Parham 1990; Schlosstein et al. 1973). There are four independent theories that may begin to explain its role in disease: arthritogenic peptide hypothesis, endoplasmic reticulum (ER) stress, cell surface free heavy chain recognition and microbial dysbiosis, with each theory having supporting and contrary evidence (Bowness 2015).

Acute and chronic gastrointestinal inflammation has been demonstrated to be present in approximately half of all AS patients (Van Praet et al. 2013), which is in addition to the previously mentioned fact that AS and CD often occur together (Mielants et al. 1995). There is evidence for the role of bacterial infection in developing reactive arthritis (Leirisalo-Repo et al. 2003), and that environmental factors as well as a genetic predisposition may be involved in the development of SpAs. This suggests that gut inflammation may be an auto-inflammatory stimulus that may drive part of the disease. HLA-B*27 may be involved in altering the gut microbiome (Rosenbaum and Davey 2011), as recent studies in HLA-B*27 transgenic rats have shown alterations in the handling of intestinal flora (Lin et al. 2014), that may permit the development of a chronic inflammatory state. Recent evidence that SpAs are governed primarily by innate immunological abnormalities suggests that the disease may occur due to a lack of protection against microbial pathogens in the gut, perhaps due to changes in gut permeability (Arrieta et al. 2006). There are links between gut and joint inflammation in patients with AS, with clinical remission being associated with normal gut histology and chronic inflammatory lesions being risk factors for the development to AS (Mielants et al. 1995). Genetic factors associated with IBD, including *NOD2* and *CARD9* (*CARD9* is also associated with AS), also participate in the response to microbial products (Zhernakova et al. 2008; Waterman et al.

2011), suggesting a role for the innate immune response. In a recent study, AS variants have shown enrichment in regulatory regions in gut cell types and gene ontology enrichment analysis identified microbes and the gut in the aetiology of AS (Li et al. 2017c), further highlighting the potential involvement of the gut microbiome in disease pathogenesis.

AS patients frequently demonstrate new bone formation and ankylosis of the spine. New bone formation is also a feature of other SpAs such as PsA, and therefore are likely to be driven by the same molecular pathways. The molecular mechanisms driving these changes are yet to be fully understood, meaning that current therapeutics such as TNF-blockers may be ineffective at reducing skeletal structural damage. Genetic associations related to radiographic severity in AS demonstrated an association with *RANK*, a gene involved in osteoclastogenesis and in the interaction between T cells and DCs and *PTGS1*, involved in the prostaglandins pathway (Cortes et al. 2015b). Drugs to retard the osteoproliferative processes in AS may assist in slowing down radiographic progression that are not targeted by current treatments. The vitamin D3 metabolite -(23S,25R)-25-hydroxyvitamin D3 26,23-peroxylactone was found down-regulated in AS patients in comparison with healthy controls, which may also contribute to skeletal changes in AS (Fischer et al. 2012).

1.3.4 Cells types involved in AS pathogenesis

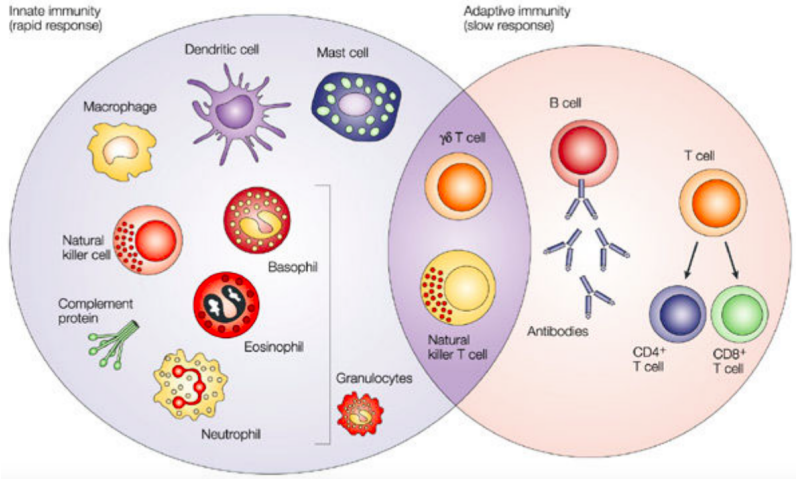


Figure 1.1: Diagram of cells in the innate and adaptive arms of the immune system. Diagram reproduced from (Dranoff 2004).

A number of cell types within the innate and adaptive arms of the immune system are believed to be important in AS pathogenesis, including monocytes, T lymphocytes and natural killer (NK) T cells (Bautista-Caro et al. 2013; Conigliaro et al. 2011; Wright et al. 2009). The innate immune system is able to trigger a rapid immune response by immediate recognition of pathogens, and the adaptive immune system is able to form long-term memory and antigen specific responses (Borghesi and Milcarek 2007). The cell types most relevant for AS pathogenesis are currently not completely described, due the limited knowledge of AS aetiology.

Certain genetic variants or single nucleotide polymorphisms (SNPs) may only exert their effects in a specific cell type, and under a specific environment such as a chronic inflammatory state. The expression of genes that specify cell function are associated with densely spaced clusters of active enhancers known as super-enhancers, that are highly cell type specific (Heinz et al. 2015). The majority of gene expression studies in AS to date have been carried out in peripheral blood mononuclear cell (PBMC) populations rather than isolated cell types, giving us a

limited understanding of how individual immune cell types may be contributing to disease (Pimentel-Santos et al. 2011; Assassi et al. 2011; Smith et al. 2008; Duan et al. 2010; Gu et al. 2009; Gu et al. 2002).

Monocytes are a crucial component of the innate immune response, which is emerging as having an important role in AS pathogenesis. Activation of monocytes results in various cellular functions such as chemotaxis, antigen presentation and the differentiation to macrophages and dendritic cells (DCs). Monocytes from AS patients have been shown to express elevated levels of proteins involved in inflammation compared with monocytes from healthy controls and also significant changes in protein expression within the ubiquitin proteasome pathway (UPP) (Wright et al. 2009). Monocyte-derived DCs revealed an altered function and gene expression pattern in HLA-B*27 axial SpA patients, with decreased stimulatory capacity of $CD4^+$ T cells (Talpin et al. 2014). IL-23, which is involved in the UPR is also present in macrophages and dendritic cells (Abraham and Cho 2009). In the inflamed mucosa, $CD14^+$ macrophages produce IL-23, TNF-α and stimulate the induction of the Th17 response.

IL-23 promotes the expansion of Th17 cells, which are a subset of $CD4^+$ effector cells recognised to be the main source of IL-17 (Langrish et al. 2005; Park et al. 2005a). IL-17, when produced in excess, is known to contribute to chronic inflammation and inflammatory diseases, therefore acts as an attractive therapeutic target (Beringer et al. 2016). A positive correlation between circulating Th17 cells and BASDAI was observed in AS (Bautista-Caro et al. 2013), and the frequencies of IL-17 positive $CD4^+$ T cells were increased in PBMCs from patients with AS and patients with RA, resulting in secretion of higher quantities of IL-17 (Shen et al. 2009). Th17 cells may be one principle driver of inflammatory disease, rather than the Th1 subset of T cells producing IFN-γ as previously thought (Shen et al. 2009).

There is a large body of evidence for the involvement of the adaptive immune system in AS. The association with *ERAP1* and HLA-B*27 antigen presentation and peptide processing demonstrated impaired $CD8^+$ T cell responses in mice (Hammer et al. 2006). The synovium and cartilage of AS patients have been shown to contain aggregates of T cells and B cells, suggestive of a role in disease aetiology (Revell and Mayston 1982; Voswinkel et al. 2001). The prevalence of Th2 and Th17 cells responsible for the regulation of adaptive immunity was higher in AS peripheral blood than in healthy controls (Szalay et al. 2012), and elevated IgA levels have also been reported in AS patients (Veys and Leare 1973; Cowling et al. 1980). B cells were shown to be involved in gastrointestinal health by producing IgA and IgM which may help to protect the epithelial barrier (Takahasi et al. 1990; Brandtzaeg 2007).

NK cells have important roles in the innate immune response by defending against microbial pathogens, but are also crucial in their ability to influence the differentiation of adaptive immune responses (Orange and Ballas 2006). The presence of certain allotypes of the killer cell immunoglobulin-like receptors (KIR) superfamily, such as KIR3DL1 (Zvyagin et al. 2010), may modulate susceptibility to AS and PsA, suggesting an important role for NK cells in the SpA pathogenesis. NK cells of SpA patients have been shown to produce Th1 cytokines, and are involved in bone damage and in the activation of dendritic cells and resident fibroblast-like synoviocyte cells (Conigliaro et al. 2011). HLA-E, the ligand for the NKG2A receptor, was found to be associated with AS in a Sardinian population, suggesting that NK activity is possible player in the pathogenesis of AS (Paladini et al. 2009). However, there are many conflicting studies about the role of NK cells in SpA pathogenesis (Park et al. 2009; Spadaro et al. 2004), driving the need for more specific and focused studies.

Neutrophils are the most abundant cell types in the immune system, where they are rapidly recruited to sites of infection (Kruger et al. 2015). They play an

essential role in proper resolution of inflammation, but inefficient regulation can trigger positive feedback loops that promote neutrophil activation and a chronic disease state. Th17 cells are able to recruit neutrophils by producing IL-17 and the neutrophil chemoattractant CXCL8, which in return recruit Th17 cells via release of the chemokines CCL20 and CCL2 (Pelletier et al. 2010). Neutrophil lymphocyte ratio (NLR) has emerged as a valuable and reliable method for follow-up of systemic inflammatory disease, and neutrophil counts were found to be significantly higher in AS patients compared with healthy controls (Mercan et al. 2016). A positive correlation between the NLR and BASDAI was observed, suggesting that the NLR may be a simple and inexpensive marker to indicate disease activity in patients with AS (Kucuk et al. 2015). Further studies are required to understand neutrophils mode of action in AS and related diseases.

1.3.5 Therapeutic interventions

Treatment for AS is currently aimed at disease management and prevention of clinical deterioration. No therapies in current use are designed to cure or reverse disease processes. Non-steroidal anti-inflammatory drugs (NSAIDs), glucocorticoids and disease-modifying anti-rheumatic drugs (DMARDs) are traditionally used in the treatment of autoimmune inflammatory diseases. However, there are various toxicities associated with these drugs (Li et al. 2017a), along with concerns that treatment with NSAIDs may cause gastrointestinal toxicity in AS patients who are at risk of an underlying bowel inflammation (Rudwaleit and Baeten 2006).

The recent discovery of anti-TNF therapy has produced 5 TNF-α inhibitors in the last two decades (infliximab, adalimumab, etanercept, golimumab, and certolizumab pegol), and has revolutionised treatment options for patients with AS (Monaco et al. 2015). These belong to the biologics class of drugs, working to regulate the immune system in a similar way to natural substances. These

differ in their molecular structure and administration, of which only infliximab is administered intravenously. Efficacy has been demonstrated for both PsA and axial SpA (Ritchlin et al. 2009; Atteno et al. 2010; Baraliakos et al. 2012; Migliore et al. 2012), and they are usually administered after NSAID failure or in the case of highly active disease (Baraliakos et al. 2012). Patients with peripheral joint involvement may also need treatment with DMARDs in case of intolerance to TNF-α inhibition (Braun et al. 2011; Ward et al. 2016). 40% of patients do not respond to anti-TNF therapies (Roda et al. 2016), and there is a lack of evidence that TNF blockers actually prevent disease progression (Anandarajah and Ritchlin 2005; Rudwaleit et al. 2004). There is no effect of anti-TNF therapy on the newly described IL-23/IL-17 axis in AS, which may begin to explain the lack of efficacy in preventing longer-term disease deterioration in a large majority of patients (Milanez et al. 2015). Side effects of TNF blockade are not uncommon, with notably increased susceptibility to infection, and can also result in the development of further inflammatory diseases, such as susceptibility to psoriasis (Askin et al. 2017), and new manifestations of IBD with etanercept use (Sandborn et al. 2001; Song et al. 2008). Overall there is a marked improvement in many clinical symptoms of AS when treated with TNF-α inhibitors (Sieper and Poddubnyy 2016). The new molecule IL-17A inhibitor secukinumab, is the first approved non-anti-TNF-α biologic for use in AS (Lubrano and Perrotta 2016), providing an alternative treatment pathway in multiple inflammatory diseases.

1.4 Genetic variation and disease

Genetic variation may contribute to a large majority of disease phenotypes. Understanding the functional consequences of this variation is important for developing novel targeted therapeutics based the molecular mechanisms of the particular disease or disease subtype in question.

1.4.1 Types of genetic variation

Genetic variation across the genome can be either single nucleotide polymorphisms (SNPs) or structural variants such as insertions and deletions (INDELs), which include copy number variants (CNVs). SNPs are the most common type of genetic variation, and can either be a substitution, insertion or a deletion (Ku et al. 2010). The relevance of a genetic variant depends on its location (Haraksingh and Snyder 2013). If located within the coding portion of the genome, a variant may lead to a change in the amino-acid sequence of the protein it codes, known as a non-synonymous mutation. Coding variants can also lead to different isoforms of the same gene due to difference in transcription start sites, protein coding DNA sequences or untranslated regions (UTRs). Coding variants with no effect are known as synonymous. The majority of genetic variation lies within the non-coding regions, either within a gene intron or between genes (intergenic).

1.4.2 Regulatory variants

Non-coding variants may have a functional effect in regulating gene expression through a variety of mechanisms. Regulatory variants have been shown to explain higher trait heritability than coding variants across 11 common diseases included in the Wellcome Trust Case Control Consortium (WTCCC) genome wide association study (GWAS) (Gusev et al. 2014). Mapping gene expression as a quantitative trait (eQTL), can reveal genetic variation associated with a change in expression, therefore linking a set of variants to a gene. eQTLs can either show association with a gene locally (*cis*-acting), that are within 1Mb on the same DNA strand, or those that demonstrate association to a distant gene (*trans*-eQTL). Large eQTL datasets in a variety of tissues are extremely useful resources in the investigation of a SNPs effect on gene expression across different cell types,

such as the Genotype-Tissue Expression (GTEx) project, reporting 44 tissues with eQTL analysis (Carithers and Moore 2015).

It has been demonstrated that variants may only affect expression under certain physiological conditions (Fairfax et al. 2014), with eQTLs varying in effect size across cell type (Fairfax and Knight 2014). To enhance our understanding of non-coding variation, large scale projects such a the Encyclopedia of DNA Elements (ENCODE) Project Consortium (Dunham 2012), the NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al. 2010) and BLUEPRINT Consortium (Adams et al. 2012), have begun to systematically annotate non-coding regulatory regions using various biochemical assays, in a large number of cell types and tissues. The identification of functional variants by overlap with regulatory annotation is particularly useful for GWAS loci with multiple disease associated variants due to linkage disequilibrium (LD). The genomic annotations can be used for global GWAS analysis to identify relevant non-coding annotations, biological pathways and causal variants (Ward and Kellis 2012a; Paul and Beck 2014).

There are many useful bioinformatic tools that enable us to determine whether a variant may have a functional potential. One such tool is RegulomeDB (Boyle et al. 2012), which includes experimental data from ENCODE and other sources, as well as computational predictions and manual annotations. These data sources are combined into a powerful tool that scores variants to help identify functional variants and their putative regulatory role. The integration of different datasets aiming to annotate the functionality of the genome can be achieved by using ChromHMM (Ernst and Kellis 2012), to systematically annotate the chromatin state based on a multivariate Hidden Markov Model. These chromatin state maps are already available for a number of cell types and tissues, enabling rapid predictions on variant functionality (see Section 1.8.4). One example of a functional follow up study is the AS associated *IL23R-IL12RB2* intergenic

region, in which epigenetic and transcription factor (TF) binding data was used to identify the probable causal SNP, of which downstream functional analysis was performed using luciferase reporter constructs, mRNA and TF binding assays. Reduced luciferase and H3K4me1 methylation was seen for the risk allele, along with reduced binding of nuclear extract, but there was however no effect on gene expression levels (Roberts et al. 2016). This may indicate that the regulatory element containing the AS associated SNP may be an enhancer for other genes, which may be determined by using chromatin conformation capture assays such as Capture-C (Hughes et al. 2014), or that there is a more complex disease mechanisms at play.

1.5 Genetics of AS

A large amount of progress has been made in the last decade in beginning to understand the genetics of AS, leading to new advances in disease pathogenesis, and the identification of possible more effective treatments.

1.5.1 Heritability of AS

The genetic component of the disease is relatively high. Family and twin studies have demonstrated that AS is highly heritable with the risk of disease in first-degree relatives of AS cases being >52 times that of unrelated subjects (Brown et al. 2000). The reoccurrence risk in monozygotic twins is 63% and in first degree relatives 8.2% (Brown et al. 2000). SpA subtypes also aggregate in families, suggesting a strong overlap between the genetic factors in these diseases, with articular and extra-articular features often segregating together (Baeten et al. 2013).

1.5.2 MHC genetics

HLA-B*27, a class I surface antigen encoded by the B locus on the MHC, is the most strongly associated genetic factor to AS in which over 100 subtypes have been identified. The association of B*2705, B*2702 and B*2704 with AS has been proven in case-control studies (MacLean et al. 1993; Lopez-Larrea et al. 1995). Associations in other subtypes still remain unclear (Garcia et al. 1998; Armas et al. 1999; Garcia-Fernandez et al. 2001; Tamouza et al. 2001). HLA-B*27 is thought to explain approximately 40% of the genetic heritability to disease, and is already implemented as a diagnostic factor (Reveille 2011a). However, 95% to 99% of HLA-B*27 positive individuals remain healthy, suggesting that other MHC or non-MHC genetic loci are also driving AS pathogenesis (Gonzalez et al. 1999). Individuals who are HLA-B*27 positive often have an association with the gene *ERAP1*, known to be involved in MHC peptide processing and presentation (Evans et al. 2011), although recent evidence has also found *ERAP1* associations in HLA-B40 individuals (Cortes et al. 2015a). Homozygosity for HLA-B*27 may increase the risk of developing AS (Jaakkola et al. 2006).

The MHC region is extremely gene dense, with many of its genes having an immunological function and the majority of immune-mediated diseases having an association with HLA. Genetic susceptibility discovery in the MHC is confounded by the complex LD structure known to exist in this region. Genetic associations with AS have been found to involve MHC associations outside of the B*27 locus, such as with HLA-E in a Sardinian population (Paladini et al. 2009). However, non HLA-B*27 studies still remain inconclusive, with one study suggesting a region spanning 270 kb of chromosome 6 and containing 23 genes, may include genes involved with susceptibility to AS (Sims et al. 2007). MHC associations can also be considered complex, involving multiple HLA and non-HLA associations, as identified by GWAS (Brown et al. 2015).

1.6 Genome Wide Association Studies

GWAS have become an important tool in the identification of novel variants that may influence the risk of developing common disease. The concept of GWAS is to analyse a set of genetic variants, most often SNPs and INDELs, using tag SNPs as markers to represent an area in the genome, due to the principle of LD and haplotype blocks (Visscher and Hill 2009). These tag SNPs are used to genotype thousands of individuals using a genotyping array, with the aim to identify genetic differences between groups of individuals such as diseased and healthy. GWAS rely on the Common Disease Common Variant (CDCV) hypothesis, which states that genetic variation with appreciable frequency in the population but relatively low penetrance in disease, are the major contributors to genetic susceptibility to common diseases (Schork et al. 2009). Most typed SNPs are relatively common, with a minor allele frequency (MAF) >5%, and therefore if associated with disease have only a modest effect with odds ratios (ORs) often between 1.1 and 2. These variants are most likely to act with an accumulation of other common risk variants. The Common Disease Rare Variant (CDRV) hypothesis, on the other hand, suggests that rare variants with high penetrance may be the major contributors to common disease due to their stronger effect on disease phenotypes, and therefore less common in the population due to its depletion by natural selection (Schork et al. 2009).

GWAS in AS have resulted in the identification of genes that would not have been identified through hypothesis driven research, such as the role of genes relating to barrier functions in the gut and the defence against microbial pathogens (Evans et al. 2011; Cortes et al. 2013). A very large number of association tests are required, and the power to detect associations will increase with sample size. Stringent criteria for significance are also required, with a *P*-value of less that 5×10^{-8} being a standard threshold. GWAS are now also beginning to include

structural variants, such as the Immunochip which contains 718 small insertion-deletions (Trynka et al. 2011; Cortes and Brown 2011).

Challenges in GWAS studies

Aside from the limited power GWAS have to detect rare variants, a number of other limitations also affect GWAS. Many of the common disease associations identified with GWAS approaches localise to the non-coding portion of the genome such gene deserts or introns. Coding SNPs represent only a small amount of the GWAS signals, often meaning the specific genes that are being regulated are unknown. The LD within the region also means that it is not possible to fine-map the exact causal variant underlying the association.

GWAS have only been able to explain a small amount of the genetic component for the majority of common diseases, leaving a large amount of the disease heritability unexplained. This is often referred to as the missing heritability, and there are a number of hypotheses to explain why this may occur (Blanco-Gomez et al. 2016; Eichler et al. 2010). One reason may be that rare and structural variants are poorly captured by existing genotyping arrays, which are are believed to have larger effect sizes in disease. Epistatic interactions (gene-gene) may also play a role, and remain largely undetected by current studies. Imprecise phenotyping and entry criteria into case-control studies may also reduce the effect sizes, and fail to capture the true genetic variants associated with the disease of interest. There may also be a difference in the genes affecting susceptibility to disease within the group, meaning that the power estimates to detect these associations may be wrong (Glazier et al. 2002; Maher 2008; McCarthy et al. 2008; Schork et al. 2009; Manolio et al. 2009).

The problem of missing heritability may be overcome by increasing sample sizes, and improving genotyping arrays to include further structural and rare variants. However it may be the case that exome sequencing and family studies may

become more important than GWAS in this endeavour (Costantino et al. 2017). Resequencing of GWAS loci may be required to identify these rare variants with a larger effect size. An example of this was performed by Rivas et al at the *CARD9* locus, where a protective rare splice variant was identified that was not described through previous GWAS studies in AS, CD or UC. Deep resequencing of 131 CD associated genes was also able to confirm three reported variants and identified eight novel variants, providing new insights into CD that would have been unavailable using GWAS alone (Hong et al. 2016).

1.6.1 Fine-mapping methods

Associated variants identified through GWAS may not be the causal SNP due to the inheritance of many closely linked SNPs in a haplotype block. Identifying the most strongly associated SNPs with disease at a locus can aid in discovery of disease mechanisms, such as how a SNP may be regulating a candidate gene. The goal of fine-mapping analysis is to identify a minimal set of SNPs that includes the causal variant(s), therefore minimising the genomic interval identified through GWAS. Fine-mapping requires the generation of a large variant set that should in theory include the causal variant, either genotyped or imputed. High density genotyping arrays such as the Immunochip (Cortes and Brown 2011) are cost effective, and the genotyping data can be imputed in publicly available large-scale sequencing panels such as 1000 Genomes Project, offering near complete coverage of variants with MAF >1% in multiple ancestry groups (Abecasis et al. 2012; Auton et al. 2015). An example of a fine-mapping analysis conducted using Immunochip data was recently carried out in AS using an approximate Bayesian approach (Wakefield 2007), leading to an improvement in causal variant identification, with seven out of twenty-six AS loci fine-mapped to less than ten variants in a credible set (Bunt et al. 2015). Bayes factors provide a measure of the strength of the association, and the credible sets generated

provide an estimate of the probability for each SNP to be causal, known as the Posterior Probability (PP), and assumes the hypothesis of only one causal variant at each locus. Conditional analysis can be performed at loci with multiple association signals, prior to credible set generation (Bunt et al. 2015). The 99% credible set represents the variants that collectively capture 99% of the PP at the locus. Comparing the credible set generation using the approximate Bayesian approach resulted in a much better refinement than by taking all SNPs in LD (e.g. r^2 >0.8), leading to a reduction in false-positives (Bunt et al. 2015).

A useful addition to fine-mapping approaches are trans-ethnic fine-mapping meta-analyses, which aim to discriminate differences in LD patterns between common SNPs in different populations (Wu et al. 2013; Mahajan et al. 2014). The haplotype diversity in the different populations was shown to improve credible set generation by using the Yoruba (YRI) component of the 1000 Genomes (Bunt et al. 2015). Overlap with disease-relevant tissue regulatory annotations can also be used to further refine the causal variant from the credible set (Vecellio et al. 2015; Roberts et al. 2016). The inclusion of genome-wide functional annotation as functional priors in the fine-mapping framework was also shown to improve fine-mapping results via the approximate Bayesian approach (Bunt et al. 2015). Another recent fine-mapping approach called probabilistic identification of causal SNPs (PICS), involved development of an algorithm to estimate the probability of a SNP to be causal given the haplotype structure and pattern of association at each locus (Farh et al. 2015). They used this approach to look at cell type specificity for the PICS SNPs by annotating the SNPs with *cis*-regulatory elements from 33 different cell types. AS PICS SNPs were found to be enriched in *cis*-regulatory elements for colonic mucosal, duodenum mucosal, monoctytes and various T cell subtypes, therefore informing follow up studies of possible relevant tissues. There is now the possibility to scale up functional experiments using massive parallel reporter assays, which would allow screening of a large

number of associated variants in a high-throughput manner (Inoue and Ahituv 2015), so that larger credible sets, or taking all candidate variants for functional follow-up would be feasible.

The generation of whole genome sequencing data may offer the most accurate analysis scheme for fine-mapping, but these datasets are currently small and therefore lack power to discriminate between variants in high LD. Resequencing of GWAS loci is one of the most accurate ways to catalogue common and rare variants (e.g. (Rivas et al. 2011)), but this method currently lack cost effectiveness due to the large sample size required to detect small effect sizes.

1.6.2 The role of GWAS data in highlighting immune-mediated disease causing loci

GWAS are a powerful tool for the investigation of disease aetiology in common complex traits, providing new insights into potential disease mechanisms. In the case of AS, there have been four major GWAS studies published, identifying at least 43 independent genetic signals (Table 1.2) which have highlighted the importance of non-MHC loci. The most recent association study is the Immunochip, a microarray SNP genotyping chip informed by previous GWAS and deep sequencing data from various autoimmune and inflammatory diseases. It demonstrated that 4.3% of AS heritability is explained by loci outside of the HLA-B region. The specific role of the Immunochip was deep replication of previous GWAS data, and fine mapping of confirmed susceptibility loci. A cross-disease meta analysis (Ellinghaus et al. 2016) has increased this number to 113 independent AS associated loci. These include genes relating to the IL-23 pathway, such as *IL23R* and *IL6R*, peptide handling and presentation involving *ERAP1*, and activation and differentiation of T lymphocytes including *RUNX3* and *ZMIZ1*. These genes are only selected by proximity to the variants, and genes unknowingly regulated at a distance are not included in the genes and

pathways described. Genome-wide studies looking at CNVs may also be able to identify further variation, such as a study by Jung et al, of which 227 CNV regions were shown to be significantly associated with the risk of AS, 9 of which were successfully replicated (Jung et al. 2014).

One important insight from GWAS is that the vast majority of associated genetic variants fall into non-coding portions of the genome, and likely impacting the regulation of gene expression. It was shown that only 12% of SNPs derived from 151 GWAS are located in protein coding regions, with 80% falling into introns or intergenic regions (Hindorff et al. 2009). This makes variants identified through GWAS difficult to interpret.

GWAS study	Sample size (AS vs Controls)	Ethnicity	Genes
(Burton 2007) WTCCC/TASC	922 vs 1466	British	*ERAP1*, *IL23R*
(Reveille 2010) TASC	2,053 vs 5,140	European N. American	2p15, 21q22, *ERAP1* *ANTXR2*, *IL23R*, *IL1R2*
(Evans et al. 2011)* TASC/WTCCC2	1,787 vs 4,800	European	*RUNX3*, *TNFRSF1A-LTBR* *CARD9*, *TBKBP1*, *PTGER4*
(Cortes et al. 2013) IGAS	10,619 vs 15,145	European Asian	*IL6R*, *FCGR2A*, *UBE2E3* *GPR35*, *ZMIZ1*, *BACH2* *SH2B2*, *NOS2*, *TYK2* *NKX2-3*, *IL27*, *ICOSLG*

Table 1.2: Summary of GWAS studies in AS (Burton 2007; Reveille 2010; Evans et al. 2011; Cortes et al. 2013). WTCCC= Wellcome Trust Case Control Consortium, TASC= The Australo-Anglo-American Spondyloarthritis Consortium, IGAS= International Genetics of Ankylosing Spondylitis Consortium. Some of the genes associated with novel findings from each study are listed. * This study demonstrated the epistatic interaction between *HLA-B*27* and *ERAP1*.

Shared autoimmune risk loci in SpA

One major role of the Immunochip was to facilitate and enable comparison of the genetic architecture between different immune mediated diseases (Parkes et al. 2013). GWAS have provided evidence for both shared and distinct genetic associations between the SpAs, highlighting a number of important pathways that may provide future opportunities for a more targeted approach to treatment. A systematic comparison between AS and PsA, for example, reveals similarities and differences, with each genetic factor bearing a relatively weak impact on the predisposition to SpA, but are indicative of shared pathways between different phenotypes (Baeten et al. 2013). Many of these shared associations may have different directional effects, such as the associated variants at locus 12p13 in AS and MS (Cortes et al. 2013; Gregory et al. 2012). The ImmunoBase web based tool enables cross comparison of associated variants with up to 20 immune mediated diseases (http://www.immunobase.org).

Ubiquitination, peptide processing and antigen presentation

ERAP1, associated with AS, psoriasis, and Crohn's disease, has been shown to interact with multiple HLA class I alleles. Epistasis (which refers to an action of a gene upon another) has been demonstrated with *ERAP1* and *HLA-B*27* and HLA-B40 in AS, and *HLA-Cw6* in psoriasis - suggesting that this interaction is critical to these diseases (Evans et al. 2011; Cortes et al. 2015b; Strange et al. 2010). ERAP1 is an aminopeptidase, responsible for trimming peptides prior to presentation by HLA class I molecules. Other aminopeptidases thought to be associated with disease are *ERAP2* and *NPEPPS* (Cortes et al. 2013). There are 3 haplotypes at the *ERAP1* locus associated with AS (Costantino et al. 2015), which are most likely to be attributable to common ERAP1 haplotypes and haplotype combinations (Roberts et al. 2017). One haplotype is tagged by the nonsynonymous SNP rs30187 which is believed to affect the enzymatic activity

of ERAP1, and another by rs10050860. There may also be SNPs affecting gene regulation of *ERAP1* transcription (Harvey et al. 2009; Costantino et al. 2015). This association suggests that the HLA-B*27 association is operating through aberrant processing of antigenic peptides (Evans et al. 2011).

Gut mucosal immunity and maintenance of barrier function

The genetic overlap between AS and IBD has been demonstrated multiple times, including the 1q32 locus and *STAT3* associations (Danoy et al. 2010), and 20 other AS associated loci also shared with the same directional effect (Parkes et al. 2013; Cortes et al. 2013; Jostins et al. 2012). These associations suggest an important role of mucosal immunity and bacterial sensing, but also a role for IL-23 signalling, as IL-23 production was seen to be increased in the terminal ileum of AS patients (Ciccia et al. 2009). Gut dysbiosis has been demonstrated in AS patients (Costello et al. 2015), suggesting that the pathogenesis of AS may be influenced by the gut microbiome, possibly by triggering IL-23 production. Genes involved in the innate response to pathogens (*CARD9*), and in bacterial sensing (*GPR35*, *GPR65* were found associated with AS (Evans et al. 2011; Cortes et al. 2013)).

IL-23 and IL-17 pathways

The IL-23 pathway is thought to be a main driver in AS pathogenesis, and the association in the *IL23R* region is among one of the most prominent genetic predispositions to SpA, shared with a number of other immune mediated diseases. The Arg381Gln polymorphisms in the *IL23R* is known to reduce risk of AS and other diseases such as CD, UC, RA and psoriasis (Abdollahi et al. 2016), decreasing sensitivity to IL-23 by reducing its efficiency in signalling through the JAK/STAT pathway (Oliver et al. 2007). IL-23 is not associated with non-radiographic forms of SpA, and SNPs in the *IL23R* may affect disease severity

rather than disease susceptibility (Costantino et al. 2017). The *IL23R* is most highly expressed on activated T cells, NK cells and more sparingly on monocytes, and its expression can be modulated by a number of factors such as IL23 and TGF-β. The IL23R is involved in $CD4^+$ T cell differentiation into Th17 cells, which are critical for host defence as well as inflammatory disease, and Th17 activation can in turn modulate *IL23R* expression.

Th17 T cells are the main producers of IL-17, and there is evidence of the IL-23/IL-17 axis role in driving intestinal inflammation (Ahern et al. 2010). Linkage at the *IL23R* and *IL17A* gene region was found to increase UC risk, suggesting the effect of a gene-gene interaction (Yu et al. 2012). Th17 cells together with IL-17^+ $CD8^+$ T cells can secrete high amounts of IL-17, contributing significantly to the cytokine production responsible for the disease microenvironment (Cosmi et al. 2008; Kleinschek et al. 2009). Several strategies targeting Th17 responses in AS have proven promising, and sucukinumab is now the first IL17A inhibitor for use in AS (Lubrano and Perrotta 2016). Ustekinumab (a monoclonal antibody against the p40 subunit of IL-12 and IL-23) resulted in a 50% improvement in BASDAI score in 55% of patients followed in an open non-randomised study for a 24 week period (Poddubnyy and Rudwaleit 2012).

TNF signalling

A number of associated variants have implicated genes involved in TNF signalling in AS, and the effective use of TNF-α inhibitors in treating AS has highlighted the role of TNF mediated pathology. *TBKBP1*, an associated gene (Cortes et al. 2013) is an adaptor protein that binds to TBK1, a kinase in the TNF signalling that plays a role in innate antiviral immunity (Fitzgerald et al. 2003), and is part of the TNF interaction network (Bouwmeester et al. 2004). TBKBP1 also interacts with the ubiquitin-dependent autophagy receptor NDP52, limiting the growth of ubiquitin-coated bacteria in the cytoplasm (Thurston et al.

2009). Two GWAS signals were identified at locus 12p13, with the primary signal associated with *TNFRSF1A*, and is in strong LD with an MS associated SNP. This SNP is a splice variant, known to result in the loss of the transmembrane domain of this receptor (Gregory et al. 2012), and can mimic TNF blocking therapies.

Intergenic regions

GWAS variants are often enriched in non-coding regions, and associations within gene deserts have been found at loci 2p15 and 21q22. The closest gene to locus 2p15 is *B3GNT2*, also shared with psoriasis (Tsoi et al. 2012). The association at locus 21q22, shared with CD and UC and a number of other diseases, is proximal to a number of gene candidates, including *PSMG1* and the TF *ETS2*. However these genes are yet to be confirmed to have a pathogenic role in AS. One unpublished study found two completely novel long non-coding RNAs (lncRNA) expressed from the 21q22 region, which were up-regulated in AS cases. Their up-regulation was demonstrated in the FANTOM5 atlas of human gene expression almost exclusively in $CD14^+$ monocytes, and may suggest a novel potential mechanism of microbial-induced $CD14^+$ monocyte expression of the 21q22 transcripts (Haynes and Thomas 2014).

1.6.3 Non-GWAS and linkage studies

Linkage studies in AS and SpAs have identified genes and polymorphisms in non-MHC loci not identified through GWAS. A meta-analysis of four whole genome linkage scans demonstrated linkage to a marker on chromosome 19q13, adjacent to the killer immunoglobulin-like receptor (KIR) genes (Carter et al. 2007). Carrying two inhibitory *KIR3DL1* alleles expressed at high levels could cause an individual to develop AS (Zvyagin et al. 2010). HLA-B*27 was shown to strongly bind to KIR3DL2, which promoted the survival of T cells in AS (Wong-Baeza et al. 2013). Further linkage scans have found an association with AS at

the cytochrome P450 2D6 gene (*CYP2D6*) on chromosome 22q13.1 (Brown et al. 2000), and strong linkage at chromosome 16q (Laval et al. 2001) and at 6q and 11q (Zhang et al. 2004). Non-parametric linkage analysis identified a new region associated with SpA at the 13q13 locus (Costantino et al. 2016). Associated polymorphisms with SpA were also described at chromosome 9q31-34 (Miceli-Richard et al. 2004) and most recently found near *MAPK14* (Costantino et al. 2017). These studies are small, and therefore require future replication.

1.7 Transcription regulation and disease

As mentioned in Section 1.4.2, *cis*-acting variants are involved in regulating gene expression and variants that disrupt TF binding motifs may lead to changes causal of disease. TFs often act as part of a complex within an enhancer or silencer site, meaning that the identification of a functional mechanism may be difficult to delineate.

1.7.1 Regulatory elements

Transcription is controlled by a multitude of regulatory elements. Due to the highly organised structure of the genome, these elements are controlled by the binding of factors at specific regulatory motifs upstream of the transcription start site (TSS) plus at local and distal regulatory elements. These regulatory elements can either be enhancers, silencers, insulators and can be part of a locus control region (LCR). The function of these regulatory elements are explained in Figure 1.2.

Multiple enhancers and enhancer variants may target the expression of a single gene, and certain genes are associated with clusters of active enhancers known as super-enhancers. The majority of super-enhancers, also known as stretch enhancers, are cell type specific, play key roles in human cell identity in health

and disease (Parker et al. 2013; Hnisz et al. 2013; Whyte et al. 2013), and offers a new role for non-coding variants in the susceptibility to common traits. Enhancers are usually regulated by the binding of TFs, and a transcription factor binding site (TFBS) may be altered by non-coding variants, ultimately affecting gene expression regulation (Bonder et al. 2017).

Enhancers can be described as inactive, primed, poised or active (Ernst and Kellis 2010). Histones at regulatory elements contain specific patterns of acetylation and methylation. For example, enhancers are often marked with H3K27ac and H3K4me1, and promoters with H3K4me3 (Creyghton et al. 2010; Heintzman et al. 2009; Heintzman et al. 2007). Histone acetylation is increased by histone acetyltransferase (HAT) and removed by histone deacetylase (HDAC), and methylation is controlled by a group of enzymes called methyltransferases. Primed enhancers are nucleosome free, but may require additional signals to become activated. Latent enhancers have also been described. These are unbound by TFs and lack the histone marks characteristic of enhancers, but can acquire these features in response to stimulation. An example of this is during macrophage stimulation. This is where deposition of enhancer marks created by the binding of stimulus activated TFs enabled a stimulus specific enhancer to be created. Many of these enhancers did not return to a latent state when stimulation ceased, enabling a faster and stronger response upon re-stimulation (Ostuni et al. 2013).

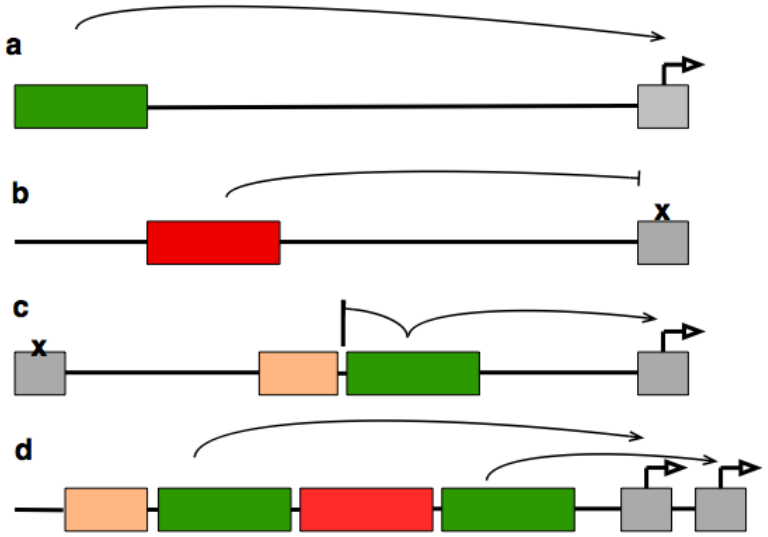


Figure 1.2: Diagram of regulatory elements and their effect on gene expression (a) Enhancer element driving gene expression downstream of the TSS by the binding of TFs. **(b)** Silencer complex inhibiting gene expression. Silencers may contain binding sites for repressors, which are negative regulatory TFs. **(c)** Enhancer element inhibited by an insulator. **(d)** Locus control region containing multiple regulatory elements, including enhancers, silencers and insulators that may be involved in regulating the expression of multiple genes. Diagram reproduced from (Maston et al. 2006).

1.7.2 Chromatin interactions

Understanding the role of regulatory elements in controlling gene expression is challenging as the genome is organised and partitioned into 3D space (Cremer et al. 2006). 3C based technologies are one strategy to identify long-range chromatin interactions (Wit and Laat 2012), and genome-wide high-throughput applications such as Hi-C (Belton et al. 2012) have been used to define interacting compartments organised into topologically associated domains (TADs). TADs represent a region of DNA within which physical interactions occur relatively frequently, and interactions rarely occur outside of these boundaries. The disruptions of TADs has been shown to lead to de novo enhancer-promoter interactions and mis-expression, demonstrating the functional importance of genome architecture (Lupianez et al. 2015). CTCF binding sites are often found

at TAD boundary elements and are associated with the control of enhancer-promoter interactions (Guo et al. 2015; Wit et al. 2015). There is increasing evidence for the role of lncRNA in controlling chromatin structure, including nucleosome positioning and chromosome looping (Bohmdorfer and Wierzbicki 2015; Shibayama et al. 2014). Chromatin conformation is highly cell specific, and may also change in response to a stimulus. For example the *IL-1* gene cluster, specifically *IL1A*, *IL1B* and *IL17F*, come into close proximity in LPS stimulated monocytes, so that upon stimulation they may be transcribed by the same transcription factory, possibly creating a LCR (Sharaf et al. 2014).

Many *cis*-acting variants act over long distances, bringing into proximity two locations such a promoter and enhancer that are far apart in linear distance. One such example are the set of obesity associated variants within *FTO*, and their long range interactions with *IRX3* (Smemo et al. 2014). A unique long enhancer domain in adipocyte progenitor cells was identified with to be controlling expression of *IRX3*, with strong haplotype-dependent regulatory activity (Claussnitzer et al. 2015).

Recent high-throughput experiments, such as Promoter Capture Hi-C have been able to assay cell type specific promoter interactions in a large number of cell types (Mifsud et al. 2015), and also the specific interactions of GWAS variants (McGovern et al. 2016). Capture-C, a high resolution 3C technology (Hughes et al. 2014), may be an important higher resolution technique to investigate disease variants involved in enhancer activity that may be influencing chromatin interactions.

1.7.3 miRNAs

MicroRNAs (miRNAs) are now known to be important in a wide variety of biological processes, and there is evidence to suggest that they have a large implication in many immune and inflammatory diseases (Sonkoly and Pivarcsi

2009). miRNAs are small single stranded non-coding RNA molecules of 18-24 bases in length, with their main function being to target 3' untranslated regions of mRNAs for their suppression or degradation (Baek et al. 2008; Guo et al. 2010). They may play an important role in the post-transcriptional regulation of up to 30% of all human genes (Sand et al. 2009). There are now 2588 mature human miRNAs in miRBase, the majority of which still lack functional validation on their mechanisms of action. However, these are believed to regulate a large number of mRNAs, with each mRNA being targeted by several miRNAs (Lim et al. 2005; Doench and Sharp 2004), and continued discovery of miRNA functions suggests that miRNAs may be an important part of almost every aspect of cell physiology.

Abnormal miRNA expression has been implicated in a number of diseases, including cancer (Lu et al. 2005; Makunin et al. 2007; Volinia et al. 2006; Alvarez-Garcia and Miska 2005), cardiovascular disorders (Care et al. 2007) and inflammatory diseases (Hansen et al. 2007; Sonkoly et al. 2007; Sonkoly et al. 2008; Kalla et al. 2014; Papanagnou et al. 2016). There have been a substantial number of studies looking at the miRNAs that are differentially expressed in AS patients versus controls, with the majority of studies in T cells and PBMCs. These studies are summarised in (Li et al. 2016). The most studied miRNA in AS is miR-146a (Chatzikyriakidou et al. 2010; Qian et al. 2016; Xu et al. 2015), due to its known importance in immunity and inflammation (Sonkoly et al. 2008). Exposing human THP1 cells to lipopolysaccharides (LPS), resulted in an increased production of miR-14a a and b (Taganov et al. 2006). miR-146a and b are regulated by LPS and NF-κB, and are also known to be implicated in a number of related inflammatory diseases, with higher expression in rheumatoid arthritis (Bogunia-Kubik et al. 2016), psoriasis (Chatzikyriakidou et al. 2010), osteoarthritis (Zhang et al. 2017), but down-regulated in SLE (Tang et al. 2009; Lofgren et al. 2012; Ji et al. 2014).

1.7.4 LncRNAs and mRNA

Another type of non-coding RNA transcript are the long non-coding RNAs (lncRNAs), which are defined as >200 nucleotides in length and are emerging as important regulators of a variety of biological processes (Mercer et al. 2009; Wang and Chang 2011). LncRNAs are involved in the differentiation of immune cell types (Sigdel et al. 2015), and are emerging as having an important role in autoimmune disease and many other complex diseases (Li et al. 2013). There have been studies looking at the association of lncRNAs and various autoimmune diseases such as RA, SLE and psoriasis (Sonkoly et al. 2005; Shi et al. 2014; Muller et al. 2014). There are now over 18,000 recognised lncRNAs in the mammalian genome, with many yet to be functionally annotated.

Since mRNA code for protein, quantifying mRNA levels can give an indication of the underlying biology. The majority of mRNAs have well defined functions, and understanding differences in mRNA levels between two conditions can give insights on important pathways and molecular mechanisms. Studies relating to mRNA and lncRNA in SpA are outlined in Chapter 3.

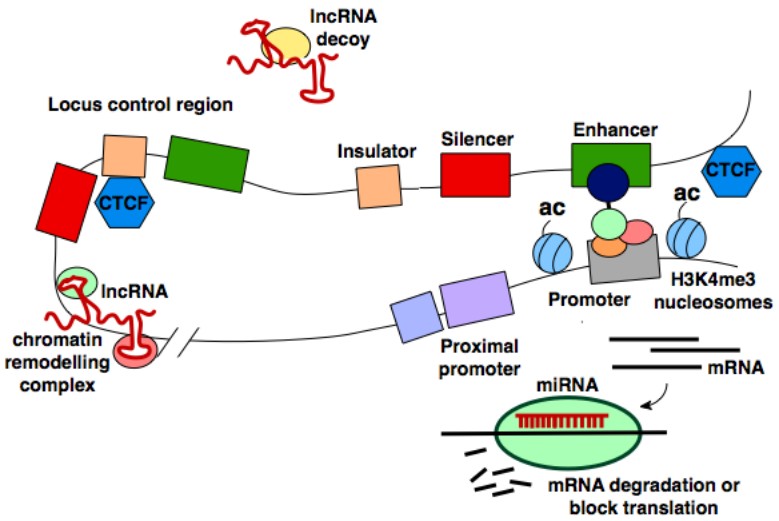


Figure 1.3: Diagram demonstrating the control of gene expression. Chromatin conformational changes may bring distal regulatory elements and promoters into proximity and allow the binding of TF complexes to regulate gene transcription. The binding of TFs also serves to maintain the chromatin in an open or active state. CTCF is involved in mediating chromatin looping and serves as a scaffold protein in regulating gene expression. Transcription is also regulated down-stream by lncRNAs, that may form part of a chromatin remodelling complex or prevent TF binding. The levels of mRNA are also regulated by miRNA, either by degradation or in preventing translation. Diagram edited and reproduced from (Maston et al. 2006).

1.8 Functional genomic approaches

While GWAS have identified disease associated variants, they fail to describe the functions that account for these observations. The post-GWAS era has begun to explore the functional relevance of how common alleles can confer risk to disease in a cell type specific manner, and also under different external environments. An example of this is the coronary artery disease associated variant at 9p21, which affects the binding of *STAT1* and lies within an enhancer for human vascular endothelial cells. Upon IFN-γ activation, the chromatin structure including STAT1 binding and the expression of neighbouring genes are altered (Harismendy et al. 2011). Therefore a key step in understanding the informativeness of GWAS for disease pathogenesis is dissecting the molecular mechanisms of disease associated variants in a cell type and context specific

manner. Functional enrichment is extremely useful in guiding fine-mapping analysis from GWAS, and can also be used to re-weight the significance of SNPs (Pickrell 2014).

There are now various techniques that make it possible to map regulatory elements of the genome, all made possible at a genome-wide scale by Next Generation Sequencing (NGS) technologies, that has massively revolutionised genomic research by increasing throughput and reducing costs. Annotations of chromatin accessibility can be defined by the hypersensitivity to DNase I cleavage (DHS) (Song and Crawford 2010) or access of the Tn5 transposase by the Assay For Transposase-Accessible Chromatin sequencing (ATAC-seq) (Buenrostro et al. 2013). These often correlate with transcription factor occupation, which can be assayed by Chromatin Immunoprecipitation of the transcription factor (TF) of choice, followed by sequencing (ChIP-seq) (Landt et al. 2012). Specific histone modifications can also be assayed by ChIP-seq, and their combinations can inform different regulatory features, which in total have been able to classify the genome into fifteen chromatin states (Ernst et al. 2011). As previously mentioned, large Consortium projects (Bernstein et al. 2010; Dunham 2012; Adams et al. 2012) are beginning to catalog these regulatory features for a large number of different cell types, and have highlighted the complexity of non-coding regulatory elements. This is useful in the effort to determine the relevant cell type and disease mechanisms underlying GWAS signals. Disease associated variants may be linked to their target gene through eQTL analysis. eQTLs are cell type specific, and only 30% of eQTLs are shared between cell types, and these tend to be located closest to the TSS (Fairfax et al. 2012). There are a number of online tools for the prioritisation and annotation of GWAS variants, such as HaploReg (Ward and Kellis 2012b), RegulomeBD (Boyle et al. 2012) and XGR (Fang et al. 2016), which all allow enrichment analysis with a variety of datasets.

1.8.1 Methods for assaying chromatin accessibility

Chromatin can be described as histone protein complexes that are associated with DNA, and can determine the accessibility of certain DNA sequences to DNA binding proteins such as TFs. Highly accessible DNA is often bound by TFs and is highly transcribed, with changes in chromatin state and accessibility tightly regulated. Changes in the chromatin structure and associated histone binding proteins are involved in cell fate determination, gene expression and can also be associated with disease (Feinberg 2007).

Until recently, chromatin accessibility has been explored by Formaldehyde-Assisted Isolation of Regulatory Elements followed by sequencing (FAIRE-seq) (Giresi et al. 2007) or DNase-I hypersensitivity followed by sequencing (DNase-seq) (Song and Crawford 2010). FAIRE-seq uses formaldehyde to cross-link DNA to proteins, and the nucleosome free DNA can then be released by sonication and phenol-chloroform extraction. DNase-seq applies DNase which preferentially digests nucleosome-free DNA and therefore will lead to an enrichement in nucleosome depleted regions. These two technologies require a large number of cells (up to 10 million cells), which is often not feasible in primary cell types. The most recent technique to investigate open chromatin is ATAC-seq, which utilises a Tn5 transposase with attached sequencing adaptors for integration into nucleosome-free open chromatin and rapid library preparation for sequencing. ATAC-seq provides similar sensitivity and specificity compared to DNase-seq, and has been able to overcome the limitations of cell number, only requiring between 500-50,000 cells, and enabling researchers to assay a larger variety of rarer cell populations that were previously unavailable. ATAC-seq is also much more time efficient, and fits into the time scale of clinical sample processing, either in the research laboratory for collection of disease cohorts, or as a future diagnostic method in the clinic (Buenrostro et al. 2013).

Genetic variation can have an effect on chromatin accessibility, due to SNPs influencing TF binding and occupancy (Maurano et al. 2015). An example of this is a minor allele variant which reduces the risk for mycardial infarction by creating a new binding motif for the Ccaat-enhancer-binding protein (C/EBP), leading to increased expression of the gene *SORT1* which lowers LDL cholesterol levels in individuals with the SNP (Musunuru et al. 2010). DNase-I hypersensitivity quantitative trait loci (dsQTL) are strongly enriched in TF binding sites, and are also associated with allele specific changes in TF binding (Degner et al. 2012), suggesting that TF occupancy may influence chromatin accessibility. There are various hypotheses that explain how a TF may act to regulate enhancer activity, such as activating chromatin remodelling and preventing nucleosome repositioning (Spitz and Furlong 2012). TF occupancy can be measured by DNase-I footprinting (Brenowitz et al. 2001), and certain TFs can be assayed genome-wide by ATAC-seq (Buenrostro et al. 2013).

1.8.2 DNA methylation

Epigenetics refers to the change in gene expression caused by modification rather than alteration of the genetic code, allowing cells that have the same genotype to exhibit a different phenotype. One such important heritable epigenetic mark is DNA methylation, which is involved in many biological processes and in the control of gene expression. DNA methylation occurs predominantly at the 5 position of cytosines adjacent to guanines defined as CpG sites via methyltransferase activity, but recent studies have also shown that the genetic code may also influence DNA methylation levels. These are known as methylation quantitative trait loci (mQTL) (McClay et al. 2015; Olsson et al. 2014), and are associated with the methylation levels of nearby CpG islands. Methylation levels are highly cell type specific (Heyn 2014; Davies et al. 2012), but it was found that sequence variants influencing methylation patterns may be

consistent across tissues (Smith et al. 2014). Complex disease aetiology may be influenced by DNA methylation, with GWAS variants influencing methylation levels (Heyn et al. 2014).

Bisulphite conversion is among the most common techniques used to study DNA methylation. Treatment of DNA with bisulphite triggers the conversion of unmethylated cytosines into uracils while 5 methyl-cytosines are unaffected. Uracils are then converted into thymine, so that the methylation levels for the cytosines can be measured by sequencing (Li and Tollefsbol 2011). Recent advances in high-throughput sequencing analysis have now allowed for epigenome-wide association studies (EWAS), and provide an opportunity to assess differences in a large number of CpGs across a cohort of patients and controls (Paul and Beck 2014). Understanding methylation changes are important, as meQTLs have been demonstrated to effect TF binding, histone modifications, and gene expression (Banovich et al. 2014).

1.8.3 TF binding and histone modifications

DNA binding proteins can be studied in a high-throughput manner by ChIP-seq, to determine the location of specific histone modifications and TF binding sites genome-wide. In ChIP, a TF or any other chromatin protein of interest is enriched by immunoprecipitation from formaldehyde cross-linked cells, along with its associated DNA. The ENCODE project has characterised TFBS in a large number of cell types (Dunham 2012), however at least one third of TF binding remains uncharacterised (Deplancke et al. 2016). There is a high degree of coordination between TF binding, nucleosome positioning and histone modifications; therefore chromatin accessibility and histone marks representative of an enhancer can act as a marker in cases where TF binding information is unavailable (Gaffney 2013). ChIP-seq has enabled the study of histone modifications and their effect on genome regulation.

1.8.4 Integration of datasets

One way to integrate all of these data types is through a software tool called chromHMM, developed by Ernst et al, which represents a multivariate Hidden Markov Model (HMM) (Ernst and Kellis 2012). It was applied to 9 chromatin marks across 9 cell types to segment the genome into 15 types of regulatory states. These regulatory states identified poised and active promoters, weak and strong enhancers, as well as transcriptionally active or inactive regions of the genome. This approach has drastically reduced the possible number of combinations of investigated chromatin marks. ChromHMM does not combine transcription factor binding information, which may be a limitation, but their classification can be useful to define regulatory potential. The Epigenome Roadmap Project investigated DNA methylation, chromatin accessibility and a large number of histone post transcriptional modifications across 111 tissues, and by using chromHMM predicted up to 50 chromatin states by combining up to 24 histone marks in the model. This approach was also useful in defining the relationship between various histone marks and methylation, with active regulatory elements being associated with H3K4me3, H3K4me1, H3K27ac, open chromatin and either low or intermediate methylation, which can be used to map GWAS variants to disease relevant cell types and tissues (Farh et al. 2015; Trynka et al. 2013).

The annotation of GWAS variants with functional data from a variety of cell types may enable prioritisation of specific variants in their specific cell type of action. Farh et al compared prioritised SNPs with H3K27ac chromatin maps for 56 cell types, and revealed the cell-type specificities of *cis*-regulatory elements for a number of immune-mediated diseases. They reported that the most relevant cell types for AS were colonic mucosa and duodenum mucosa cells, followed by T helper cells including Th17 and monocytes (Farh et al. 2015). A more

recent study reported similar results with AS associated variants enriched in enhancer-related epigenetic regions in monocytes, $CD4^+$ and $CD8^+$ T cells, NK cells, regulatory T cells and B cells and mucosa from the small intestine, sigmoid colon and rectum (Li et al. 2017c).

1.8.5 Methods for assaying gene expression

Profiling the expression of genes can begin to give insight in the genes that contribute to molecular mechanisms, and therefore is an extremely useful tool in understanding the pathogenesis of diseases. Differential expression between healthy and diseased samples can produce gene signatures of disease in terms of up or down regulated genes. Microarray analysis is the most utilised technique so far for target identification of messenger RNA (mRNA) and is based on oligonucleotide probes hybridising complimentary DNA (cDNA) from mRNA, but have a number of limitations such as background noise and low sensitivity (Hal et al. 2000). Microarrays do not allow for detection of novel transcripts, low abundance transcripts or splice variants.

RNA sequencing (RNA-seq) is a next generation sequencing method which is now widely adopted for genome-wide measurement of transcript abundance, and has several advantages over microarrays (Wang et al. 2009). These are mainly related to its increased sensitivity and low background noise, since single end or paired end RNA reads are mapped to unique regions of the genome. It can also be used to detect novel transcripts, and certain library preparation methods allow for the detection of coding and non-coding RNAs, such as long (>200 base pairs) non-coding RNAs (lncRNAs) and microRNAs (miRNAs) which are short non-coding RNAs between 18 and 24 bases in length. RNA-seq requires a relatively small amount of input RNA, and can be used to detect alternative splicing events and allele specific expression, among a large number of other analyses (Wang et al. 2009; Mortazavi et al. 2008).

There are a number of choices to make when planning an RNA-seq experiment, such as: the number of replicates, library preparation methods and sequencing depth. There are a variety of enrichment methods available to ensure that the right class of RNA is selected for sequencing, such as: removal of ribosomal RNA to ensure that all mRNAs and lncRNAs are conserved, enrichment of polyadenylated RNA for protein coding genes, or size selection enrichment for miRNAs. RNA is then converted to cDNA, fragmented and adaptors are ligated for library amplification. Multiplexing is achieved by the addition of unique indexes for sequencing on a platform of choice, such as the Illumina HiSeq series. Certain analyses such as detecting alternatively spliced transcripts may also require higher sequencing depth (Liu et al. 2013). The number of biological replicates to achieve accurate differential analysis results is difficult, especially in disease cohorts that may differ phenotypically. It has been demonstrated that the number of differentially expressed genes plateaus at four replicates (Zhang et al. 2014b), and technical replicates are not necessary due to the high reproducibility that can be achieved (Marioni et al. 2008).

1.8.6 Functional validation approaches

The techniques explained previously do not provide direct proof of a causal variant or enhancer function. Therefore it is crucial to validate the causality and mechanism of action of predicted enhancers. Plasmid based reporter assays have been widely used to characterise putative regulatory regions, and there have been several high-throughput techniques developed to simultaneous assay hundreds of thousands of reporter plasmids at once (Dailey 2015; White 2015). A general concern about these reporter assays is that they may not accurately reflect the function of enhancer elements in their native context. The validation of predicted functional variants has come a long way with the use of genome editing using the clustered regularly-interspaced short palindromic repeats (CRISPR)/Cas9

system (Cong et al. 2013; Mali et al. 2013), allowing a predicted enhancer to be assayed in its true genomic and cellular environment. CRISPR based studies can be used to assess the function of large genomic regions, where single guide RNAs can pinpoint sequences to be tested (Sanjana et al. 2016; Diao et al. 2016; Korkmaz et al. 2016). Issues with transfection difficulties in certain cell types can be overcome by using pluripotent stem cells, followed by differentiating these edited cells into the cell type of choice (Xue et al. 2016). An example of this is with the identification of Parkinson's disease associated risk variant in a non-coding distal enhancer element that regulates the expression of α-synuclein (SNCA), by combining GWAS data, genome-wide epigenetic information and (CRISPR)/Cas9 genome editing in human pluripotent stem cells (Soldner et al. 2016).

1.8.7 Gene Ontology and pathway analysis

Gene Ontology (GO) is an initiative for the annotation and curation of genes and gene products, with the aim to enable functional interpretation of experimental data via enrichment analysis. The Gene Ontology Consortium (GOC) has been developing terms to describe each gene product with the goal of data collation across various data sources (http://www.geneontology.org). These include a full set of terms for immunological processes, which will now be described as Immune Ontology (IO). GO can be used to identify statistically enriched pathways or functional gene groups (Dimmer et al. 2008) with the aim to discover gene data sets to guide development of hypotheses and to explain, for example, the alterations in response to certain diseases (Ishikawa et al. 2009). Sets of genes may point to pathways that were previously not recognised as important in a trait, and therefore may enable the identification of novel sets of biomarkers and targets for research, with the aim of novel target discovery for therapeutics (Gassen et al. 2008). Unexpected links are formed between immune associated

genes and non immune associated pathways (Gassen et al. 2008), and this cross talk will enable the improvement in our understanding of disease mechanisms.

One such tool enabling the interpretation of a molecular profile is Gene Set Enrichment Analysis (GSEA), which is a computational method that determines whether a defined set of genes is statistically significant between two biological states (Subramanian et al. 2005). The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA, of which 17779 gene sets are divided into 8 major collections and several sub-collections. Hallmark gene sets summarise and represent specific well-defined biological states or processes, generated by a computational methodology based on identifying overlaps between gene sets in other MSigDB collections and retaining genes that display coordinate expression (Subramanian et al. 2005). There are also other gene sets that give more specific molecular and cellular enrichment, such as the Canonical, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome and BioCarta pathways, which have recently been integrated into one tool to enable fast and efficient enrichment analysis across databases (Fang et al. 2016).

1.9 Specific aims and objectives of this thesis

1. To investigate differential gene expression in AS patients in specific immune cell types. (Chapter 3). Defining the cell specific differences in gene expression in AS patients may provide insights into disease aetiology that may not have been previously identified in PBMC based studies. This work will explore

- cell specific changes in gene expression in 6 immune cell types in AS.
- the relationship between gene expression and disease activity.
- the effect of B*27 genotype in healthy controls.

2. To define the chromatin accessibility landscape in AS patients, and AS patients treated with TNF-α inhibitors. (Chapter 4). Identifying regions in the genome that are differentially accessible in AS patients may lead to novel understanding of disease mechanisms. This work aims to investigate

- the differences in chromatin accessibility in AS patients and healthy controls, along with post anti-TNF therapy changes in 3 primary immune cell types.
- how changes in chromatin accessibility link to changes in gene expression in the same individuals.

3. Perform fine-mapping analysis with the aim to prioritise AS GWAS SNPs for future follow up, and perform functional characterisation. (Chapter 5). GWAS are not able to identify the causal SNP involved in disease aetiology. Therefore statistical methods can be employed in order to prioritise SNPs for follow up. This chapter will aim to

- fine-map Immunochip AS GWAS SNPs using genotyping data to obtain a final list of prioritised SNPs at each locus.
- integrate data with patient specific ATAC-seq, to identify potential molecular mechanisms.
- describe initial functional characterisation of a GWAS associated locus.

Chapter 2

Materials and Methods

2.1 Sample recruitment

All samples collected are part of the GENExpress ID study, established in 2013 to investigate the functional genomics of inflammatory disease.

2.1.1 AS patient recruitment

Patient blood samples were collected in collaboration with the research nurses at the Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Trust and Professor Paul Bowness at the Botnar Research Centre, University of Oxford. Up to one hundred millilitres of blood was collected from eligible patients with written informed consent based on the GENExpress ID study criteria in anticoagulant EDTA-containing blood tubes (Vacutainer System, Becton Dickson). All patients excluding GNORLAS019 had at least one copy of HLA-B*2705, which was determined by genotyping coding variants rs1140412, rs1071652 and rs41556417 using the RNA-seq data which for B*27 positive individuals should code an Asparagine at position 97 of the protein (Cortes et al. 2015b).

Eligibility criteria for recruitment were AS patients:

- over 18 years old.

- newly diagnosed and in flare, with no steroid medication or beginning biologic therapy for the first time.
- able to provide clinical information and written consent.
- haemoglobin of >9g/dL.
- less than 2 weeks without antibiotics unless used for prophylaxis.
- fulfil the clinically accepted criteria set out for diagnosis using the Assessment of SpondyloArthritis international Society (ASAS) classification criteria for axial spondyloarthritis (axSpA) (Lipton and Deodhar 2012).

2.1.2 Healthy volunteer recruitment through the Oxford Biobank

Healthy volunteers specifically for the RNA-seq pilot study were recruited from the Oxford Biobank, OCDEM, University of Oxford with the help of Jane Cheeseman (Oxford Biobank research nurse) and Matthew Neville (Oxford Biobank co-ordinator). Individuals were recruited based on their HLA-B*27 haplotypes, and were either homozygous for HLA-B*2705, homozygous for HLA-B*0801 (a common HLA-B haplotype), or heterozygous. The individuals recruited are shown in Chapter 3.

2.2 Sample processing

Blood samples were immediately processed, and the PBMCs and PMNs were separated by Ficoll-Paque density gradients. PBMCs and PMNs were washed twice in Hank's balanced salt solution without calcium or magnesium (Thermo Fisher Scientific), and were resuspended in PBS with 0.5% FCS and 2mM EDTA in preperation for the isolation of cell specific cell types. Cells were counted manually using a haemocytometer.

2.2.1 Primary cell isolation

Primary cells were isolated following two separate immune cell isolation work flows, depending on whether 3 or 6 cell types were being isolated (Figure 2.1). Due to time constraints and logistics of sample processing, cell populations were isolated by magnetic beads rather than by Fluorescence-activated cell sorting (FACS). Cell separations were carried out using magnetic-activated cell sorting (MACS, Miltenyi). Positive separations were used for $CD19^+$ B cells, $CD4^+$ and $CD8^+$ T cells, $CD14^+$ monocytes and $CD16^+$ neutrophils, and negative selection for natural killer (NK) cells following the manufacturer's instructions. Isolating cells by manual counting instead of by FACS means that the number of cells are not as exact for downstream applications.

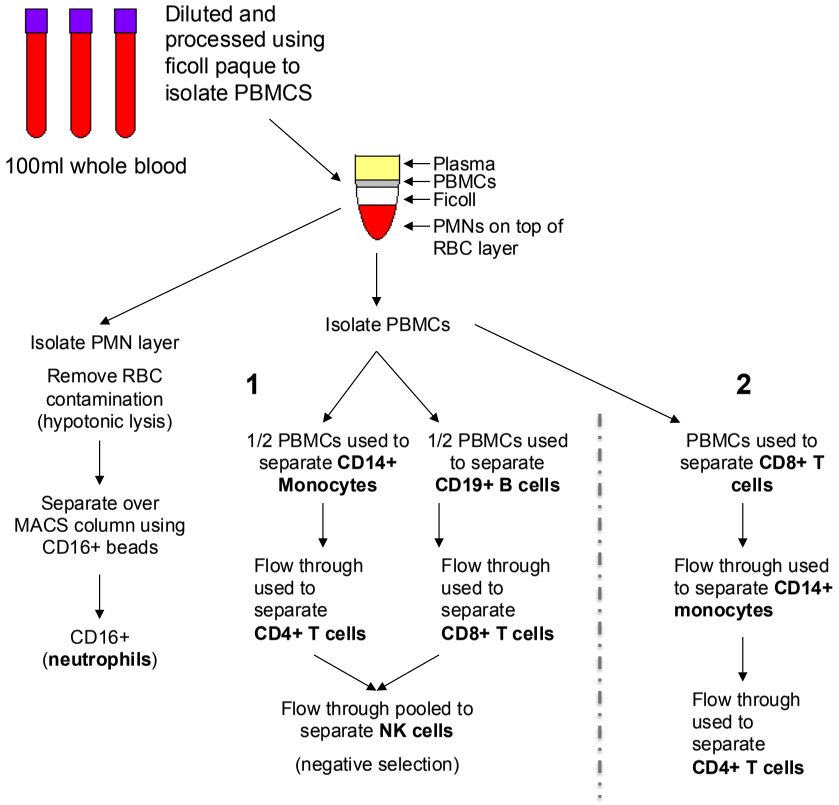


Figure 2.1: Blood processing and cell isolation workflow Method 1 demonstrates the process for 6 cell types, including $CD16^+$ neutrophils, $CD14^+$ monocytes, $CD19^+$ B cells, $CD4^+$ and $CD8^+$ T cells and natural killer cells. Method 2 describes purification of $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells only.

2.3 Experimental protocols

2.3.1 ATAC-seq

Primary cells, specifically $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells once isolated were immediately prepared for ATAC-seq. Cells were counted manually using a haemocytometer and resuspended in PBS with 0.5% FCS. ATAC-seq is advantageous over other similar protocols as very small numbers of cells are required to investigate open chromatin. A diagram demonstrating the technique and ATAC-seq workflow in detail is shown in Chapter 4, Figure 4.1.

Since manual cell counting is not as accurate as FACS sorting which was the method used in the original protocol (Buenrostro et al. 2013), ATAC-seq in my experiments was performed on between 50,000 to 100,000 cells as described (Buenrostro et al. 2013; Buenrostro et al. 2015). Cell lysis was carried out for 10 minutes, and the DNA transposition for 40 minutes using a Nextera Tn5 transposase (Illumina) and purified using the PCR Minelute Kit (Qiagen). Libraries were prepared using Nextera Primers supplied by Buenrostro et al for 11 cycles, and purified using the PCR Minelute Kit (Qiagen). To ensure the removal of adaptors and primer dimers, another round of purification was carried out using Agencourt AMPure XP Magentic Beads (Beckman Coulter). Libraries were then quantified using a library quantification kit (Kapa Biosystems) in accordance with instructions provided by the supplier. The libraries were submitted to the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics for sequencing on the HiSeq4000 (Illumina) at roughly 12 samples per lane.

2.3.2 RNA extraction and RNA-seq

Total RNA was extracted in batches from isolated primary cells previously stored at -80°C using two different extraction methods depending on the cohort. For cohort 1 (Chapter 3), total RNA and microRNAs of each sample were isolated according to manufacturers protocol of TRIzol Reagent (Life Technologies, Inc.) and mirVana microRNA Isolation Kit (Ambion, Inc.). We decided to use TRIzol for extraction of the second round of control sample recruitment, so that the two batches were processed in the same way to minimise differences. One batch of the RNA extractions was carried out by Dr Andrew Brown. Library preparation and indexing were performed by Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics, either with the ribosomal RNA depletion method (Ribo-Zero rRNA Removal Kit, Illumina) for mRNA and lncRNAs and small RNA-seq library preparation (Illumina) for miRNAs. Ribosomal RNA depletion was the preferred method over selecting for polyadenylated transcripts due to the possibility to also analyse the information for non-polyadenylated transcripts. These may include all nascent pre-mRNA (unspliced), and a number of other functionally relevant lncRNAs. However, due to the additional RNA species, up to twice as much sequencing depth is required to reach the same level of detection sensitivity for expression analysis. Small RNAs such as miRNAs are not able to be detected by simply ribo-depleting the sample, therefore to generate good quality sequencing data of small RNA it is necessary to employ a separate protocol.

It is important to note that certain technical issues may introduce variability in RNASeq data, which relate to RNA preparation, the generation of cDNA, library preparation, and the sequencing reactions themselves (Li et al. 2014a). These biases were minimised by being consistent with RNA isolation protocols, and requesting the generation of cDNA and sequencing libraries for all samples to be carried out together. This minimises the external factors, such as laboratory

conditions that may introduce bias. All samples were randomised and sequenced together, using the same sequencing chemistry.

For cohort 2 (Chapter 4) cells were homogonised using the QIAshredder (Qiagen) and RNA was isolated with the RNeasy Mini Kit (Qiagen) instead of the TRIzol Reagent and mirVana microRNA Isolation Kit used for the previous set of samples. We decided to update our extraction protocol for this batch of RNA-seq data since it can be analysed as a stand alone experiment if necessary and also due to the difficulty and dangers of phenol-chloroform extractions.

RNA was quantified with NanoDrop (Thermo Scientific), but further quantification and quality control was carried out by Core Genomics at The Wellcome Trust Centre for Human Genetics before final libraries were made for sequencing, including Bioanalyzer (Agilent Technologies) profiles and RIN to assess RNA quality and degradation. RNA was also more accurately quantified using The Quant-iT PicoGreen assay (Thermo Fischer Scientific). Input concentration of RNA for the Ribo-Zero library preparation (Illumina) was then decided based on the most suitable amount for all samples. 750ng was chosen for input material for all samples for Ribo-Zero library preparation. This method is able to capture coding RNA plus multiple forms of long non-coding RNA by depleting ribosomal RNA, and then synthesising cDNA. A small subset of the samples were also prepared for small RNA-seq (Illumina) to capture miRNAs in $CD14^+$ monocytes, and $CD4^+$ and $CD8^+$ T cells. All libraries were sequenced on the HiSeq4000 (Illumina).

2.3.3 Functional genomic assays

Luciferase reporter gene assay

DNA fragments surrounding the SNP of interest were generated by PCR amplification of genomic DNA from healthy volunteers and PGF cell line. PCR

primers for enhancer sites were designed with restriction sites BamHI/ SalI downstream of the luciferase gene, and for the *CARD9* promoter XhoI/HindIII upstream of the luciferase gene in the Promega pGL3prom vector (Figure 2.2). Constructs were confirmed with Sanger sequencing. HEK293T cells were transfected using Lipofectamine LTX and PLUS reagent (Invitrogen). Cells were left to recover for 2 hours after transfection, and if required were stimulated with IFN-γ for 24 hours. Cells were lysed and collected after 24 hours and luciferase assays were performed, normalised against pRL-TK Renilla (Promega) to correct for transfection efficiencies. Enhancer constructs were also normalised against a stuffer region predicted not to exhibit any enhancer activity.

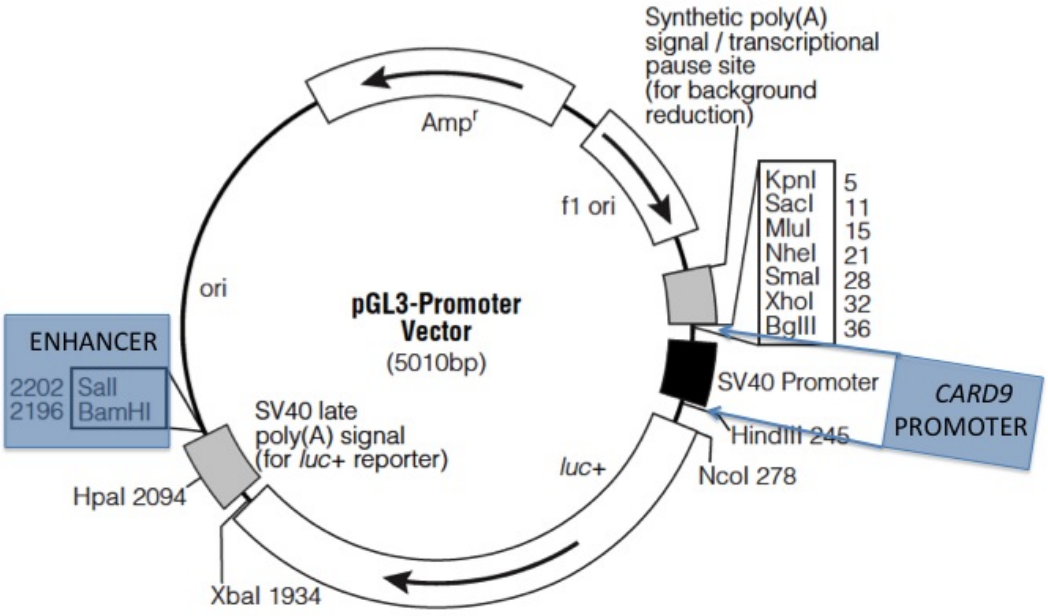


Figure 2.2: pGL3 promoter vector. Diagram of promoter and enhancer insertion sites.

Electrophoretic mobility shift assays

Nuclear extracts were prepared from primary monocytes, either unstimulated or stimulated with IFN-γ at a concentration of 20 ng/ml for 24 hours, as previously

described (Schreiber et al. 1989). Oligonucleotide probes were radiolabelled with 32P dCTP (Perkin-Elmer), and EMSA performed as previously described (Udalova et al. 1998). Probes were generated by annealing forward and reverse oligonucleotides at 15pmol. Supershift assays were performed with antibodies to YY1 (sc1703), USF1 (C-20) and c-Myc (sc764) (Santa Cruz).

2.4 Bioinformatic analysis

2.4.1 RNA-seq data processing

Total RNA-seq

RNA-seq data generated using the ribo-depletion protocol was mapped against the comprehensive annotation file for hg19 from Gencode using Tophat2 (Kim et al. 2013), containing the reference chromosomes, scaffolds, assembly patches and alternate loci (haplotypes). These comprised a total of 2840283 gene annotations including long non-coding RNAs. Reads were then filtered using Picard tools (Broad Institute) and Bamtools (Barnett et al. 2011) to remove duplicated and non-properly paired reads. Sorted and unique reads were carried forward as input to generate gene counts using HTSeq-count (Anders et al. 2014). Differential gene expression analysis was carried out using DESeq2 (Love et al. 2014), which begins by filtering out any genes with counts less than one to reduce multiple testing. Covariates were included into the model when necessary, accounting for batch effects and sex.

miRNA

Small RNA-seq reads were trimmed and then mapped against the genome build hg19 using Bowtie (Langmead et al. 2009) by Core Genomics. HTSeq-count (Anders et al. 2014) was used to identify and count mappings to known miRNA that are listed in miRBase (currently 1881 miRNA sequences) (Ambros et al.

2003; Kozomara and Griffiths-Jones 2014). Differential gene expression analysis was carried out using DESeq2 (Love et al. 2014), and those with less than one count were filtered out to reduce multiple testing. Covariates were included in the model, accounting for disease activity and sex. miRNA gene targets were selected using only high confidence experimentally validated interactions using miRTarBase (Chou et al. 2016), (http://mirtarbase.mbc.nctu.edu.tw/index.php). These miRNA targets were used to overlap with differentially expressed mRNAs.

Pathway analysis

Pathway enrichment analysis was carried out using XGR (Fang et al. 2016), specifying a background of all genes included in the differential expression analysis. Gene set enrichment analysis was carried out using Hallmark gene sets, Canonical, KEGG (Kanehisa and Goto 2000), REACTOME (Fabregat et al. 2016) and BioCarta Pathways all from the Molecular Signatures Database (MSigDB) Collections (Subramanian et al. 2005) using a hypergeometric distribution and removing redundant terms.

For visualisation, KEGG pathways were manually curated to include all genes of gene families to enable overlap between gene symbols and expression data. An R function was created by Hai Fang to enable colouring of the pathway based on fold change and genes passing significance highlighted in bold. All immune related pathways will eventually be manually curated to form part of an R package and web-app for Immune Ontology and the integrative Analysis using a manual curated Atlas of pathways (A2).

2.4.2 ATAC-seq data processing

ATAC-seq data was processed by an in house pipeline that I contributed towards, which performs individual sample data processing, along with creating a merged master list for each comparison for differential open chromatin analysis.

For each individual sample, adapter sequences and low quality bases from fastq files were trimmed using cutadapt (Martin 2011). Bowtie2 (Langmead et al. 2009) was used to align the reads to the genome build hg19. Picard tools (Broad Institute) was used to mark duplicate reads and Samtools (Li et al. 2009) was used to remove duplicated and non-properly paired reads. Peak calling was performed using MACS2 (Zhang et al. 2008), using shift-100, extension size 200 to align the peak with the cut site. These parameters are standard for DHS and ATAC-seq data analysis. For overlapping peaks, the mean of each peak summit is calculated and then replaced by a new summit, and then extended +/-250 base pairs to create the new peak. Peaks at $<10^{-8}$ were used for downstream analysis. BigWig files were produced for visualisation using bedGraphToBigWig (ENCODE). For each sample, a quality control metric to test the Irreproducibility Discovery Rate (IDR) between pseudo-replicates was used to ensure that the peak summits have been built in the same way for the entire sample peak list. Bad quality samples were filtered out if this IDR percentage was below 45%. Enrichment around transcription start sites (TSS) is also a way to measure for quality in ATAC-seq samples, since we expect the TSS across the genome to be enriched in comparison with flanking DNA. Plotting this was also a way to filter out bad quality samples.

A master list for the comparison of interest is built containing non overlapping 500bp peaks present in at least two samples of all the samples to be compared. The counts are then retrieved at each of these positions by HTSeq-count (Anders et al. 2014) to obtain a count file for differential analysis. DESeq2 (Love et al. 2014) was used to detect differences in read counts at each position, with date of sequencing used as a covariate in the model to account for batch effects of library preparation. Each peak was annotated by two genes in its vicinity by Zinba (Rashid et al. 2011). The genes at each peak were intersected with differentially expressed genes from the same individuals in the RNA-seq analysis, to identify

biologically relevant overlap. This list was also used to overlap with reported AS GWAS genes.

2.4.3 Fine mapping

AS associated loci identified by GWAS are represented by a 'lead' SNP, defined by the strongest signal of association at that locus. Many lead SNPs are not primarily associated with disease susceptibility but are nevertheless associated by virtue of being in LD with the disease SNP. It may also be the case that there are multiple causal SNPs at a single locus. To prioritise GWAS SNPs for functional follow up, fine mapping was performed on a subset of the data from the International Genetics of Ankylosing Spondylitis Consortium (IGAS) (Cortes et al. 2013). Any distinct association signals, resulting from multiple independent causal variants at a single locus were identified statistically through conditional analysis. Conditional analysis adjusts for the lead SNP in the model, to test for independently associated SNPs.

GWAS dataset

To improve the identification of potential causal variants for AS, comprehensive fine mapping was carried out in 24 of the reported non-MHC susceptibility loci from the IGAS study. I was given access to a subset of the European cohort, more specifically the UK cohort, which includes 4101 cases and 9700 controls which have been genotyped with the Immunochip - a chip designed in 2009 by investigators of 11 distinct autoimmune and inflammatory diseases (Trynka et al. 2011; Cortes and Brown 2011). Quality control of the data set had already been performed, as well as assessment of the origins of the samples by PCA (Cortes et al. 2013).

Imputation of genotypes

A 2Mb genomic region surrounding the lead Immunochip GWAS SNP was subset using PLINK 1.9 (https://.cog-genomics.org/plink/1.9/) (Chang et al. 2015). Genotypes were prephased with SHAPEIT (Delaneau et al. 2011), and missing genotypes were imputed using IMPUTE2 (Howie et al. 2009), using the 1000 Genome Project Phase 3 Version 3 reference panel (October 2015 release). Imputation quality was assessed and filtered using QCtool, and those that were directly typed and well imputed in at least 70% of the sample size were carried forward for association analysis. Filtering these SNPs is important in the credible set generation, as the quality of the imputation will now have a minimal effect on calculating the posterior probability. The posterior probability is determined by effect size, which can be directly affected by the quality of the imputation.

Association testing and conditional analysis

Association analysis and conditional analysis was carried out using SNPTEST with a Bayesian additive model. The Wellcome Trust Case Control Consortium (WTCCC) previously used SNPTEST in their analysis of 7 GWAS (Burton 2007). For each association signal in the locus, credible sets of SNPs were constructed containing SNPs most likely to be driving the association signal, along with their posterior probability. The 50% and 95% credible set was generated for each signal, and SNPs were ranked according to their approximate Bayes factor, including further variants until their cumulative posterior probabilities exceeded 0.99. The original publication contains further details on the methodology (Bunt et al. 2015).

Chapter 3

Differential gene expression in Ankylosing Spondylitis

3.1 Introduction

3.1.1 Cell specificity in disease

Expression studies from mixed cell populations, such as PBMCs, lack the resolution to detect specific changes within cell subsets. PBMCs contain a wide variety of cells from the innate and adaptive arms of the immune system, which if analysed together may be unable to untangle the specific changes related to the function of one cell type. PBMC samples may vary in cell type number and composition, biasing gene expression studies (Wong et al. 2016). The majority of expression studies are now focusing their efforts in isolating specific cell subsets of interest for a more accurate representation of gene expression related to biological function, such as the large consortium projects including Roadmap, ENCODE and BLUEPRINT (Bernstein et al. 2010; Dunham 2012; Adams et al. 2012).

Understanding cell specific gene expression in a disease context can aid in the discovery of functional regulatory variation. The importance of purified cell populations has been demonstrated in eQTL studies (Fairfax et al. 2012), and also context specificity such as an immune stimulus (Fairfax et al. 2014).

Understanding the transcriptional landscape in disease across cell subsets may highlight important cell types in disease aetiology, their enriched pathways and molecular mechanisms. Cell type specific expression can be linked to predicted enhancers, and may inform the function of genetic associations identified through GWAS (Edwards et al. 2013). As described in Chapter 1, AS associated variants were found to be enriched in enhancer regions in colonic mucosa and duodenium mucosa cells, together with T helper cells including Th17 and monocytes (Farh et al. 2015), and in a more recent study in monocytes, $CD4^+$ and $CD8^+$ T cells, NK cells, regulatory T cells and B cells and mucosa from the small intestine, sigmoid colon and rectum (Li et al. 2017c).

3.1.2 mRNA expression datasets in AS

There have been a number of gene expression studies in AS, but the majority are of small sample size and based in a mixed cell population most commonly involving PBMCs (Table 3.1). Gene expression studies in SpA and AS have been unable to provide reproducible results, due to the low sample sizes, and have also not found much overlap with GWAS genes. This may either highlight the problem of missing heritability in GWAS studies, or demonstrate patient heterogeneity within both AS GWAS and gene expression studies. It may also be the case that cell populations are too mixed to achieve a meaningful assessment of differential expression, or that the wrong cell types and tissues are being assayed. Some of the main findings and pathways that have been highlighted as differentially expressed in AS and one SpA study in PBMCs and macrophages are briefly summarised here (and Table 3.1):

- Antigen processing and presentation related genes (Fang et al. 2015).
- Downstream TGF-β genes *SMAD3*, *RUNX3*, *ZAP-70*, *PPP2R1A* and *SHC1* (Chen et al. 2013).

- Innate immune genes such as TLR4 and TLR5 (Assassi et al. 2011).
- T cell differentiation (Pimentel-Santos et al. 2011).
- Reduced expression of immune related genes STAT3/JAK2 which are downstream from IL-23 in Th17 cells (Duan et al. 2010).
- Downregulation of *TNFAIP3*, *NR4A2* and *CD69* in AS patients (Duan et al. 2010).
- Innate immune components SPARC, SLPI, and NLRP2 in the pathogenesis of SpA (Sharma et al. 2009).
- IFN-γ dysregulation in macrophages (Ma et al. 2013; Smith et al. 2008).
- CXCR4/SDF-1 as a potential pro-inflammatory axis (Gu et al. 2002).

Table 3.1: Gene expression studies in AS All current published AS gene expression studies, including papers with reanalysis of previous data in GEO archive. *Cohort 1 AS n=21, uSpA n=28, Cohort 2 AS n=23, uSpA n=18, Controls n=20.

Study	Year	Cell type	Sample size	Reference
GSE41038 Bead Chip	2013	Synovial biopsies	AS=2, Co=4	(Thomas et al. 2013)
GSE25101 Microarray	2011	PBMCs	AS=16, Co=16	(Pimentel-Santos et al. 2011)
Bead Chip	2011	PBMCs	AS=16, Co=14	(Assassi et al. 2011)
GSE11886 Microarray, RT-qPCR	2008	PBMCs	AS=8, Co=9	(Smith et al. 2008)
Bead Chip	2009	Whole Blood	SpA=18, Co=25	(Sharma et al. 2009)
Microarray and RT-PCR	2009	PBMCs	AS=18, Co=18	(Duan et al. 2010)
Microarray and RT-qPCR	2009	PBMCs	2 cohorts AS & uSpA*	(Gu et al. 2009)
Microarray	2002	PBMCs	AS=7, Co=7	(Gu et al. 2002)
Reanalysed				
GSE41038, GSE25101, GSE11886	2015	Macrophages, Synovial, Wh.blood		(Fang et al. 2015)
GSE11886 microarray	2013	PBMCs	AS=8, Co=9	(Ma et al. 2013)
GSE25101 Microarray	2013	PBMCs	AS=16, Co=16	(Chen et al. 2013)

LncRNAs in AS

Currently there are only two reported studies looking specifically at lncRNAs and AS. One study is a microarray in AS osteogenically differentiated mesenchymal stem cells (MSC) versus healthy controls (Xie et al. 2016), which found differences in TGF-β signalling and specifically 4 lncRNAs (*ZNF354A-1*, *LIN54-1*, *FRG2C-3*, and *USP50-2*). The second study investigated the role of AK001085 and its influences on the diagnosis of AS (Li et al. 2017b). AK001085 was found to be significantly down-regulated in AS patients, and was proposed as a potential diagnostic biomarker. The tissue specific roles of lncRNAs in disease have been extensively documented (Kumar et al. 2013; Wapinski and Chang 2011) along with their involvement in microbial susceptibility (Gomez et al. 2013), and in the regulation of the immune system, such as the innate immune response (Peng et al. 2010) and $CD4^+$ and $CD8^+$ T cell differentiation (Collier et al. 2012; Pang et al. 2009). The role of lincRNA-cox2 has been demonstrated in NF-κB signaling (Carpenter et al. 2013), and increased levels have been shown to be associated with rheumatoid arthritis. LncRNAs found in autoimmune disease associated loci were shown to be expressed 2.5-fold more than all lncRNAs, suggesting a strong involvement of lncRNAs in disease pathogenesis (Hrdlickova et al. 2014). Studies in autoimmune diseases have found differences in the expression of lncRNAs in comparison to healthy controls, such as the up-regulation of *NEAT1* in SLE (Zhang et al. 2016) and *NEAT1* and *TUG1* in RA (Song et al. 2015) patients.

3.1.3 miRNAs in AS

There is growing evidence to suggest an important role for miRNAs in the pathogenesis of AS, as well as many other inflammatory diseases, as previously described in Chapter 1. Studies of miRNAs in AS have identified a small number of miRNAs with potential importance in AS (Li et al. 2016), but for the majority there is a lack of understanding in their mechanism of action. Three studies

analysed genotype associations at *miR-146a* (Chatzikyriakidou et al. 2010; Qian et al. 2016; Xu et al. 2015); a well studied miRNA due to the known importance of *miR-146a* in the immune system (Sonkoly et al. 2008). Two of these studies found a significant association with AS, one with rs3027898 on *miR-146a* target *IRAK1* (Chatzikyriakidou et al. 2010), and also with *miR-146a* polymorphism rs2910164 (Xu et al. 2015). Rs3027898 is in high LD with rs1059703 which has been previously found to increase NF-κB activation, and in turn activate *miRNA-146a* (Taganov et al. 2006; Arcaroli et al. 2006). The first miRNA expression profile study was carried out in AS T cells, which discovered up-regulation of 8 miRNAs and down-regulation of 5 miRNAs in AS T cells. *Let-7i*, *miR-221* and *miR-16* were also found to be correlated with BASRI (Lai et al. 2013). *MiR29a* was also found to be elevated in the PBMCs of AS patients compared with rheumatoid arthritis patients or healthy controls (Huang et al. 2014). These finding have pointed to potential regulation of the cytokine response, T-cell survival and the regulation of bone loss (Li et al. 2016).

3.2 Aims

This results chapter aims to determine differences in gene expression between AS patients and healthy controls. Previous studies in AS have focused on gene expression analysis in mixed cell populations such as PBMCs, and also more specifically in macrophages and T cells. The pilot study reported in this chapter aims to look at specific cell populations with the aim of understanding the cell type specificity of protein coding genes, lncRNAs and miRNAs that may be deregulated in AS patients versus healthy controls. This is the first study that will look simultaneously at up to 6 primary immune cell types from the same individual believed to have a role in AS development. The relationship with disease activity will also be explored. The evidence for possible involvement of the selected cell types in AS disease aetiology is described in Chapter 1.

Specific aims are:

1. To compare levels of expression of mRNAs and lncRNAs between AS patients and healthy controls in six cell types and perform gene set enrichment analysis to determine important pathways involved in disease for each cell type.

2. To establish miRNAs differentially expressed between AS patients and healthy controls and integrate this with mRNA data.

3. To investigate whether there are differences in global gene expression related to possession of *HLA-B*27*.

3.3 Results

3.3.1 Patient and control cohorts

Table 3.2 details the 10 AS patients recruited for the pilot study of gene expression in AS analysed by total RNA-seq in six cell types through the GENExpress ID study. The six patients with high disease activity (GNORLAS004, GNORLAS014, GNORLAS017, GNORLAS018, GNORLAS019, GNORLAS035) have an average BASDAI score of 6.8 (range 4.6-9.3) and a mean age of 26.6 years and 8.7 years post AS diagnosis. The 4 patients with low disease activity (GNORLAS036, GNORLAS037, GNORLAS038, GNORLAS041) have an average BASDAI score of 1.7 (range 0.7-2.7) and a mean age of 32.5 years and 5.9 years post diagnosis. These samples were also used for small RNA-seq for miRNAs but only in 3 cell types. All AS patients except GNORLAS019 had at least one copy of *HLA-B*27*. Further details of the BASDAI classification for each patient can be found in Chapter 7, Table 7.1.

Healthy volunteer sample collection included two rounds of recruitment from the Oxford Biobank (OBB), due to quality issues from the first round of

recruitment. The aim was to recruit healthy volunteers enriched for different *HLA-B*27* genetic backgrounds, in order to identify *HLA-B*27* specific effects in healthy volunteers (Table 3.3).

For 3 cell types ($CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells), patients GNOPBAS043-GNOPBAS047 from the second cohort recruitment were also included for mRNA and lncRNA to increase sample numbers. Chapter 4 (Table 4.1) describes these patients in further detail.

Patient ID	Sex	Age at Diagnosis	Disease duration (months)	BASDAI
GNORLAS004	Male	27	48	8.2
GNORLAS014	Female	27	48	4
GNORLAS017	Female	38	242	5.7
GNORLAS018	Male	27	6	9.3
GNORLAS019	Male	18	156	8
GNORLAS035	Male	23	72	6.3
GNORLAS036	Female	35	108	0.2
GNORLAS037	Female	42	108	0.5
GNORLAS038	Male	21	48	1.3
GNORLAS041	Male	32	120	0

Table 3.2: AS cohort 1 recruitment. The BASDAI or Bath Ankylosing Spondylitis Disease Activity Index consists of a 1 - 10 scale measuring discomfort, pain, and fatigue (1 being no problem and 10 being the worst problem) in response to six questions asked of the patient pertaining to the five major symptoms of AS. Patients GNORLAS004 to GNORLAS035 are described as having high disease activity, whereas GNORLAS036 to GNORLAS041 have lower disease activity. All patients are of Caucasian origin. These samples were previously recruited by the research nurses and former group members.

Patient ID	Sex	Age	HLA-B27 genotype
Recruitment 1			
OBB2657	Female	34	*2705/*0801
OBB3073	Female	54	*2705/*0801
OBB4267	Female	46	*2705/*0801
OBB4633	Female	37	*2705/*0801
Recruitment 2			
ASHVC06	Male	40	*2705/*0801
ASHVC07	Female	42	*0801/*0801
ASHVC08	Male	55	*0801/*0801
ASHVC10	Female	41	*2705/*0801
ASHVC11	Female	44	*2705/*2705
ASHVC12	Female	54	*0801/*0801
ASHVC13	Female	51	*2705/*2705
ASHVC14	Male	47	*0801/*0801
ASHVC15	Female	52	*2705/*2705
ASHVC16	Male	54	*2705/*0801
ASHVC17	Female	49	*2705/*0801
ASHVC18	Female	51	*0801/*0801
ASHVC19	Female	43	*2705/*0801
ASHVC20	Female	45	*2705/*0801
ASHVC21	Female	36	*2705/*2705
ASHVC22	Male	42	*2705/*0801
ASHVC23	Female	37	*2705/*2705

Table 3.3: Oxford Biobank Recruitment. Samples from the first round of recruitment that passed quality control (collected by a previous member of the Knight Lab), along with the new cohort recruited by myself in 2017.

Patient characteristics

The patients were assessed by questionnaire on a number of questions relating to their lifestyle and health including smoking status, family history of AS, other autoimmune diseases and extra articular manifestations (Table 3.4).

Table 3.4: AS cohort 1 metadata PJI=Peripheral Joint Involvement

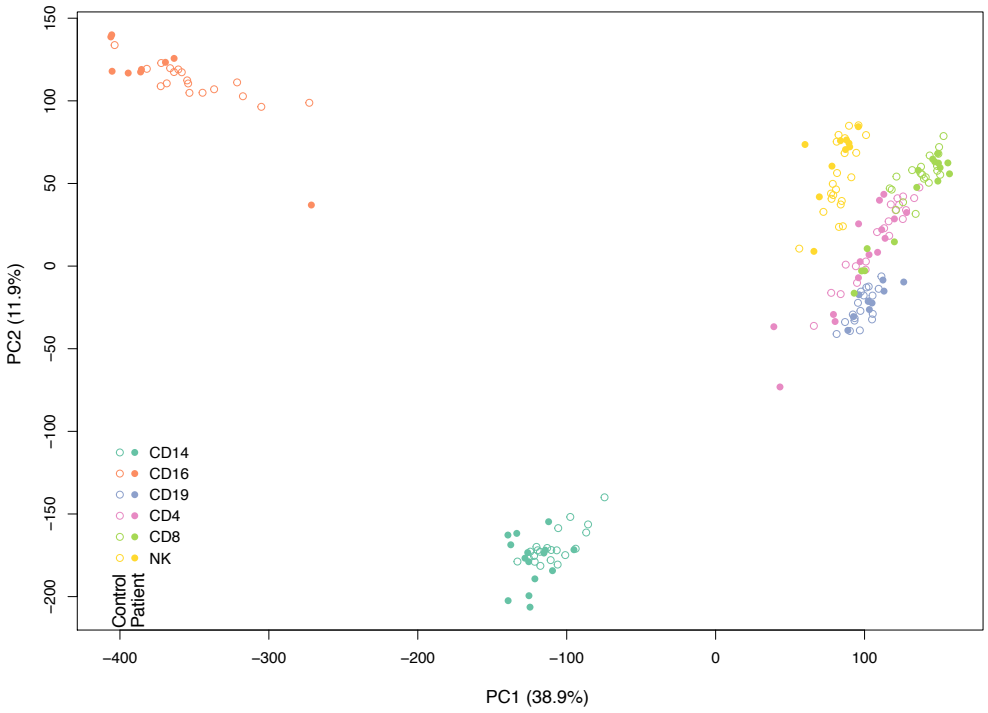
Patient ID	Smoker	Cigarettes/Day	Family history of AS	Other Autoimmune	Extra-articular
GNORLAS004	Current	5	No	Parent with Crohn's disease	PJI
GNORLAS014	Current	10	Great Grandmother	No	Uveitis
GNORLAS017	Previous	1	No	Sibling with Psoriasis	Uveitis
GNORLAS018	Current	4	Sibling	No	None
GNORLAS019	Previous	25	Yes	Don't know	PJI
GNORLAS035	No	-	Father	Grandmother with Hashimoto	Uveitis
GNORLAS036	Previous	20	Aunt	Don't know	PJI and Uveitis
GNORLAS037	No	-	No	Sibling with Psoriasis	PJI and Uveitis
GNORLAS038	No	-	Father	Yes	Uveitis
GNORLAS041	No	-	Father	No	None

3.3.2 Differential expression of mRNAs and lncRNAs between AS patients and controls

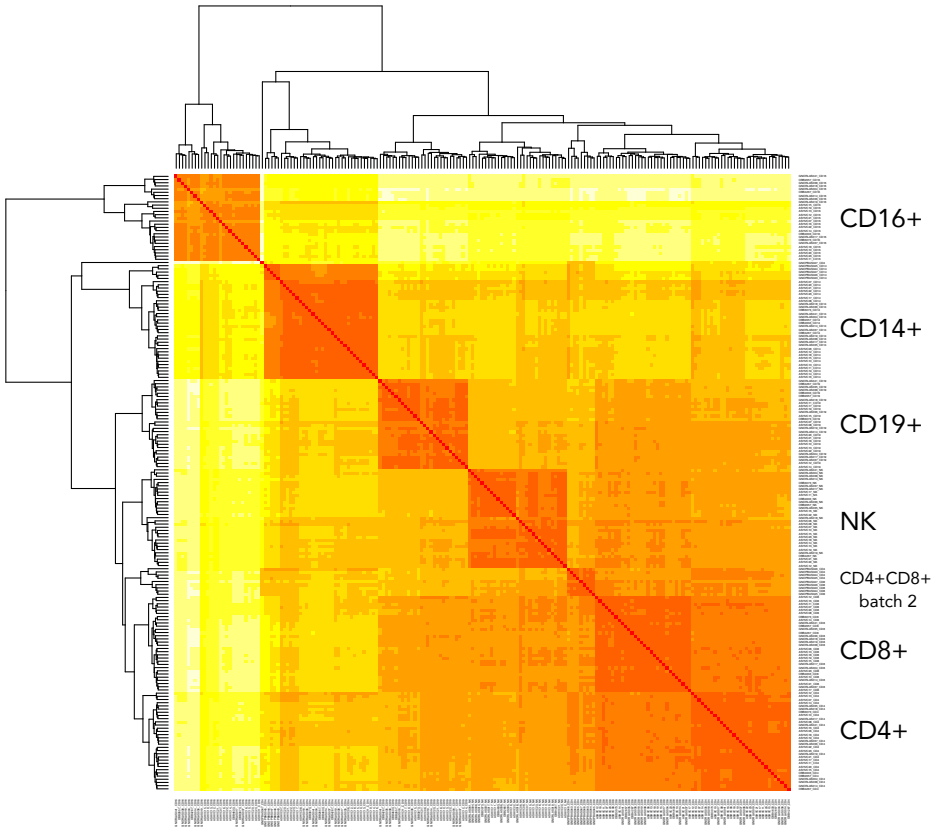
As previously described, the Ribo-Zero rRNA Removal Kit (Illumina) was used to remove ribosomal RNA and enrich for coding mRNA transcripts and lncRNAs. AS samples from cohort 2 (Chapter 4) were also included in this analysis to increase numbers in 3 of the cell types.

Quality control

AS and healthy control samples were mapped using Tophat2 and gene counts generated using HTSeq-count as described in Chapter 2. The number of counts mapping to a comprehensive list of mRNA and lncRNAs exceeded 10 million reads per sample, with only 4 samples having slightly lower read numbers (approximately 8 million reads) but were randomly distributed among the samples and therefore kept for downstream analysis. The AS samples from Cohort 2 were also included for $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells to increase the number to 15 AS samples (except for $CD4^+$ in which one sample is excluded (see Chapter 7, Figure 7.2.1). PCA of the first two PCs is shown in Figure 3.1(a), demonstrating clear clusters between cell types, and a seperation between AS patients and controls. Unsupervised hierarchical clustering illustrates the relationship between clusters of samples based on reads which map to genes. The height of the branches indicates the strength of separation between individuals. The separation between cell types is demonstrated clearly in the distance matrix heatmap in Figure 3.1(b), however the AS samples from cohort 2 create their own cluster, especially in $CD4^+$ and $CD8^+$ T cells. This will be accounted for during the differential expression analysis.



(a)



(b)

Figure 3.1: QC of total RNA-seq samples. **(a)** PCA **(b)** Hierarchical Clustering.

Differential expression mRNA

Differential expression analysis was carried out using DESeq2 (Love et al. 2014), correcting for the batch effect in the 3 cell types with the additional AS samples from AS cohort 2. The model also corrected for sex, since the AS samples were not sex matched to healthy controls. Disease activity was not taken into account in this pilot analysis. The number of differentially expressed genes are listed in Table 3.5, filtered by a False Discovery Rate (FDR) value of either <0.05 or <0.01. The largest number of differentially expressed genes were identified for $CD16^+$ neutrophils, which is consistent with a previous study that discovered substantially increased transcriptional variability in neutrophils compared to monocytes and T cells (Ecker et al. 2017). Volcano plots demonstrating the differentially up and down-regulated genes with an FDR <0.05 and the log2 fold change values are shown in Figure 3.2.

Cell Type	mRNA	lncRNA
	FDR 0.05/0.01	FDR 0.05/0.01
$CD14^+$	1887/611	266/98
$CD4^+$	205/76	32/14
$CD8^+$	166/54	21/8
$CD16^+$	3375/1657	647/331
$CD19^+$	1216/450	141/48
NK	581/277	120/66

Table 3.5: The number of differentially expressed RNAs and lncRNAs between healthy controls and AS patients. An FDR cut off of <0.05 and 0.01 was used to define differentially expressed genes. Downstream analysis was carried out using the cut off of <0.01.

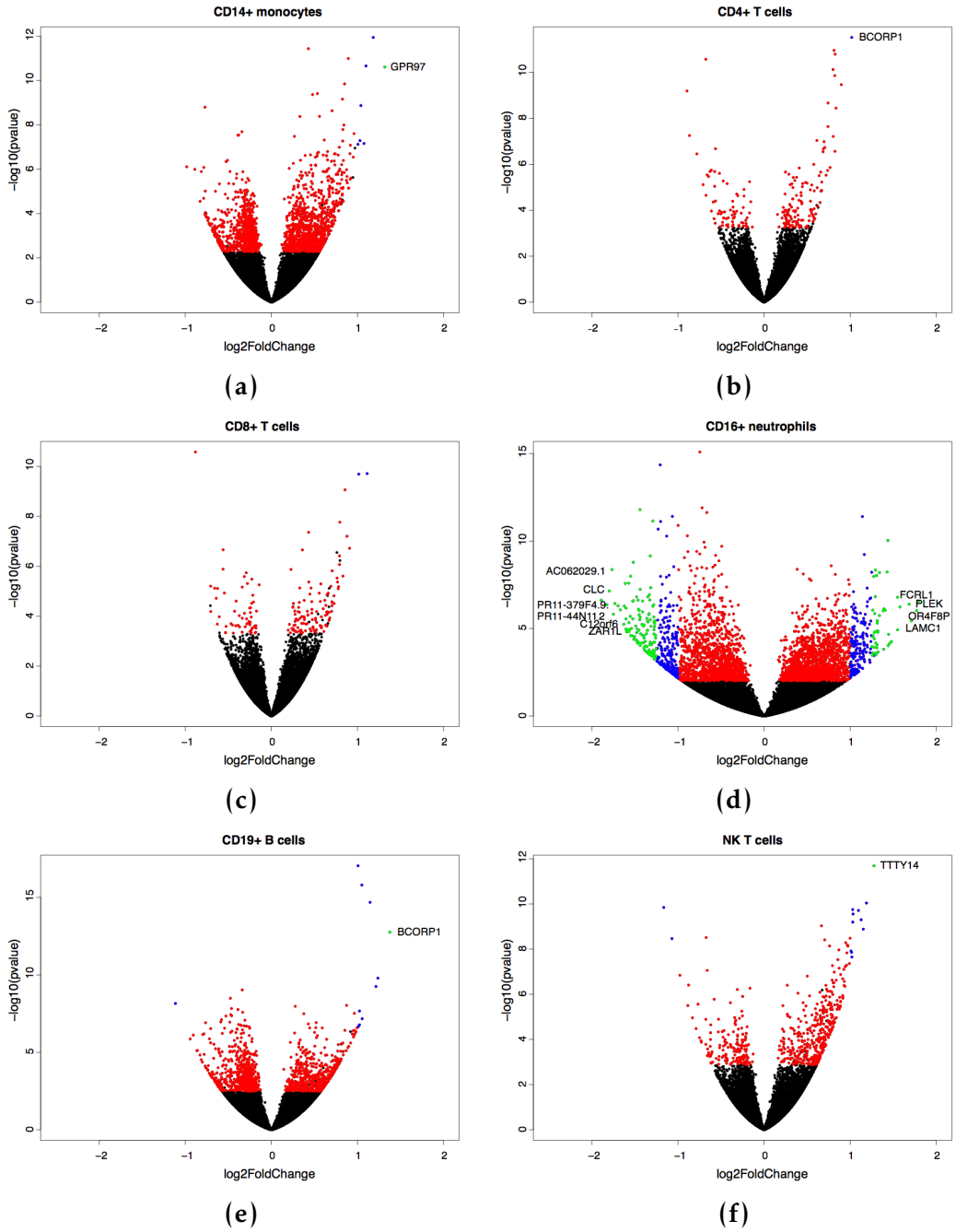


Figure 3.2: Volcano plots of differentially expressed genes in AS patients versus healthy controls. The red points demonstrate genes DE with an FDR <0.05, blue points with a Log2FoldChange >1 and green points with a Log2FoldChange >1.25 in **(a)** $CD14^+$ monocytes **(b)** $CD4^+$ T cells **(c)** $CD8^+$ T cells **(d)** $CD16^+$ neutrophils **(e)** $CD19^+$ B cells **(f)** NK T cells.

Overlap with GWAS reported genes

A list of reported GWAS genes was retrieved from the The NHGRI-EBI Catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas),

followed by manual curation to ensure that all genes from recent studies were included in the list. The differentially expressed genes for each cell type at an FDR <0.05 were overlapped with this list (Table 3.6). $CD16^+$ neutrophils demonstrated the highest overlap of 11 genes, followed by $CD19^+$ B cells with 8 genes. There were a number of overlapping differentially expressed GWAS genes across cell types, including the down-regulation of *PTGER4* in $CD4^+$, $CD8^+$ T cells and $CD19^+$ B cells. *UBE2L3* was found down-regulated in CD14+ monocytes, $CD19^+$ B cells and NK T cells, and *NPEPPS* up-regulated in CD14+ monocytes, $CD16^+$ neutrophils and NK T cells. *IL1R2* was found up-regulated in $CD4^+$, and *IL1R1* up-regulated in $CD16^+$ neutrophils and NK T cells.

Cell Type	mRNA up-regulated	mRNA down-regulated
$CD14^+$	*ZMIZ1*, *NPEPPS*, *CMC1*	*B3GNT2*, *IKZF3*, *UBE2L3*, *TBX21*
$CD4^+$	*IL1R2*	*PTGER4*
$CD8^+$	*STMN3*	*PTGER4*
$CD16^+$	*RBMX*, *FNBP1*, *IL1R1*, *NPEPPS*	*TNFRSF1A*, *KIF21B*
	IL6R, *FCGR2A*, *NFKB1*	*TYK2*, *GPR65*
$CD19^+$	*GSDMB*, *SULT1A1*, *FGFR1OP*	*DAG1*, *PTGER4*, *RUNX3*
		IKZF3, *UBE2L3*
NK	*IL1R1*, *NPEPPS*	*UBE2L3*

Table 3.6: Differentially expressed reported GWAS gene in 6 cell types. The up or down regulation refers to the expression levels in the AS patients compared to healthy controls. FDR <0.05.

Pathway analysis based on differentially expressed mRNAs

Pathway enrichment analysis using XGR (Fang et al. 2016) was carried out for each cell type using differentially expressed genes (FDR <0.01) to ensure highly relevant pathways were selected. The gene sets from Canonical, KEGG, REACTOME and BioCarta (Subramanian et al. 2005) were used, and all genes

from the differential expression analysis were used as a background set of genes. Each cell type demonstrated enrichment for distinct pathways, but there was also a number of overlapping pathways across cell types.

$CD14^+$ monocytes pathway enrichment analysis

Pathways enriched in $CD14^+$ monocytes and the differentially expressed genes (FDR <0.01) are show in Table3.7. $CD14^+$ monocytes were the only cell type to be enriched for genes involved in Antigen Presentation. *HLA-B* was not found to be differentially expressed, but instead MHC class I gene *HLA-C* and MHC Class II genes *HLA-DOA*, *HLA-DPA1* and *HLA-DRA* were found to be significantly down-regulated in patients. ER genes *PDIA3* and *CALR* were found significantly down regulated in patients, along with the proteosome subunit gene *PSME1*. There was also enrichment for the canonical NF-κB pathway and for signalling events mediated by HDAC Class I molecules, which will be discussed in more detail later in this section.

Pathway	Genes
Metabolism of RNA	*PRKCA, PABPC1, PSMB7, PSME1, PSMD7, RPL14 RPL13, RPS3, RPS4Y1, RPL18,RPL7A, RPS19 RPS12, RPL36, RPL38, LSM4, LSM1, LSM5 CNOT6, SNRPB, SNRPD1, SNRPD2, SNRPG ZFP36L1, SMG1, SMG7*
Translation	*PABPC1, SEC61G, SEC11C, SRP9P1, RPL14, RPL7A RPS19, RPL13, RPS3, RPS4Y1, RPL18, RPS12 RPL36, RPL38, EIF3C, EIF3H, EIF3I*
mRNA Splicing	*POLR2A, POLR2C, SF3B5, SNRPB SNRPD1, SNRPD2, SNRPG, SNRNP40*
Neuropeptides VIP and PACAP inhibit the apoptosis of activated T cells	*NFKBIA, CALM1, PPP3CB, EGR3, EGR2*
Regulation of Telomerase	*SP1, MAX, NCL, XRCC6, SAP30 SAP18, RBBP7, TERF2IP*
Telomeres, Telomerase, Cellular Aging and Immortality	*PRKCA, POLR2A, XRCC6, TEP1*
Signaling events mediated by HDAC Class I	*NFKBIA, NCOR2, MAX, SAP30, SAP18 GATAD2B, RBBP7, TFCP2*
Antigen Presentation: Folding, assembly and peptide loading of class I MHC	*CALR, HLA-C, PDIA3, SEC13*
ER-Phagosome pathway	*CALR, PSMB7, HLA-C, PDIA3, PSME1, PSMD7 SEC61G*
Canonical NF-kappaB	*NFKBIA, PRKCA, TNFAIP3, ERC1*
Spliceosome	*LSM4, LSM5, SF3B5, BUD31, SNRPB SNRPD1, SNRPD2, SNRPG, SNRNP40, PCBP1*

Table 3.7: Pathways enriched in $CD14^+$ monocytes between AS patients and healthy controls using differentially expressed genes FDR <0.01.

$CD16^+$ neutrophils pathway enrichment analysis

Enrichment for canonical NF-κB pathway and signaling events mediated by HDAC Class I molecules were also found enriched in neutrophils, showing a trend of these pathways towards innate immune responses in AS. All significantly enriched pathways and genes at an FDR <0.01 are shown in Table 3.8.

Pathway	Genes
Spliceosome	*DDX5, HSPA2, SNW1, LSM3, TRA2B, RBM8A EIF4A3, RBM22, HNRNPC, WBP11, SYF2, SF3B5 BUD31, SNRPA1, SNRPB, SNRPC, CWC15PHF5A, DDX42 SNRPG, ALYREF, PCBP1, PRPF38A, SMNDC1, BCAS2 NCBP2, SRSF2, SRSF5, SRSF6, SRSF7*
SRP-dependent cotranslational protein targeting to membrane	*SRP14, SPCS3, RPL41, RPL9, RPS3 RPL17, RPS26, RPS27, RPS29, RPL23, RPL26 RPL30, RPL27A, RPS10, RPL34, RPL35A, RPL37 RPL38, RPS15A, TRAM1, SSR2, SSR4*
Signaling events mediated by HDAC Class I	*NFKBIA, CREBBP, HDAC7, HDAC4, TNF TNFRSF1A, PRKACA, SUMO1, SSPO, SAP18 GATAD2B, WDR77, RBBP7, TFCP2, GATAD2A*
RNA Polymerase II Transcription	*CDK7, TAF4, TAF5, GTF2A2, TAF9, TAF13 POLR2K, RBM8A, SNRPB, SNRPG, ALYREF, NCBP2 SRSF2, SRSF5, SRSF6, SRSF7, CLP1, PABPN1, CCNT1*
Translation	*EIF5, EIF1AX, SRP14, SPCS3, RPL41, RPL9 RPS3, RPL17, RPS26, RPS27, RPS29, RPL23 RPL26, RPL30, RPL27A, RPS10, RPL34, RPL35A, RPL37, RPL38, RPS15A, EIF3I, TRAM1, SSR2, SSR4*
Canonical NF-kappaB	*NFKBIA, CHUK, TNFAIP3, TNF, TNFRSF1A, RIPK2, SSPO*
IL8- and CXCR2-mediated signaling events	*GNB1, PPP2CA, PDPK1, HCK, CBL, RAB7A PLCB2, ELMO1*

Table 3.8: Pathways enriched in $CD16^{+}$ neurtrophils between AS patients and healthy controls using differentially expressed genes FDR <0.01.

$CD19^+$ B cells pathway enrichment analysis

$CD19^+$ B cells had the largest number of significantly enriched pathways, with the top significant pathway being the mTOR signaling pathway. Not much is known about the role of mTOR in B cells, but it may have important roles in metabolic programming and antibody class switching (Jones and Pearce 2017). The canonical NF-κB pathway was again shown to be enriched in B cells. Significantly enriched pathways and genes (FDR <0.01) are shown in Table 3.9.

Pathway	Genes
mTOR signaling	*MAPK1, RAF1, YWHAB, YWHAZ, EEF2 MAPKAP1, YY1, RRAGC, POLDIP3*
Fc Epsilon Receptor I Signaling in Mast Cells	*PAK2, MAPK1, RAF1, PPP3CA NFATC1, LYN*
Prolonged ERK activation	*MAPK1, RAF1, YWHAB, FRS2*
BCR signaling	*NFKBIA, MAPK1, RAF1, PPP3CA NFATC1, LYN, BCL10*
Canonical NF-kappaB	*NFKBIA, TNFAIP3, RIPK2, BCL10*
Mechanism of Gene Regulation via PPARα	*NFKBIA, SP1, MAPK1, MED1 DUSP1, NCOA1*
CD40L Signaling	*NFKBIA, TNFAIP3, DUSP1*
S1P4 pathway	*MAPK1, GNA12, GNA13*
Rap1 signalling	*RAF1, YWHAB, YWHAZ*
Regulation of cytoplasmic and nuclear SMAD2/3 signaling	*MAPK1, NUP153, TGFBRAP1*
Calcineurin-regulated NFAT-dependent transcription in lymphocytes	*NFATC1, JUNB DGKA, IRF4, IKZF1*
AP-1 transcription factor network	*SP1, TP53, NFATC1, DUSP1, JUNB, BAG1*
Adaptive Immune System	*NFKBIA, PAK2, RAF1, LYN, SH3KBP1, YWHAB YWHAZ, RIPK2, VHL, RAB7A, TAB2, HLA-C CD74, BCL10, ANAPC4, ANAPC5, AP2B, PSMA8 UBE2Z, KIF20A, ORAI1, MAPKAP1, ASB6*
GMCSF-mediated signaling events	*MAPK1, RAF1, LYN, YWHAZ*
Class I PI3K signaling events mediated by Akt	*RAF1, YWHAB, YWHAZ, MAPKAP1*

Table 3.9: Pathways enriched in CD19$^+$ B cells between AS patients and healthy controls using differentially expressed genes FDR <0.01.

$CD4^+$ and $CD8^+$ T cells pathway enrichment analysis

T cell receptor signaling pathways were found enriched in both $CD4^+$ and $CD8^+$ T cells, potentially demonstrating the downstream effects of dysregulated antigen presentation pathways. *LAT* was up-regulated in both types of T cells; the role of its gene product is to act as a linker for the activation of T cells. *ZAP70* was upregulated in $CD4^+$ T cells only, and functions together with the Src family kinases in the initial step of TCR-mediated signal transduction. All pathways found enriched at an FDR <0.01 are listed in 3.10.

$CD8^+$ T cells showed enrichment for IL-1 and TNF-α mediated activation of the NF-κB pathway through TAK1, with down-regulation of *RIPK2* and *NFKBIA*, and the up-regulation of *AGER*, its protein product known as RAGE. Significantly enriched pathways at FDR <0.01 and selected pathways of interest at FDR <0.05 are listed in Table 3.11).

Pathway	Genes
T Cell Signal Transduction	*NFKBIA*, *LAT*, *TONSL*
Validated targets of C-MYC transcriptional activation	*MTA1*, *MMP9*, *PMAIP1*
GPCR downstream signaling	*CXCR2*, *DGKA*, *HCAR3*

Table 3.10: Pathways enriched in $CD4^+$ T cells between AS patients and healthy controls using differentially expressed genes FDR <0.01.

Pathway	Genes
TAK1 activates NF-κB by phosphorylation and activation of IKKs complex	*NFKBIA*, *RIPK2*, *AGER*
Adaptive Immune System	*NFKBIA*, *RIPK2*, *HLA-C*
Immune System	*NFKBIA*, *RIPK2*, *HLA-C*, *AGER*
Signaling events mediated by HDAC Class I*	*NFKBIA*, *HDAC7*, *HDAC10*, *YY1*
TCR signaling*	*NFKBIA*, *RIPK2*, *LAT*
Signaling by NOTCH*	*HDAC7*, *HDAC10*, *RAB6A*

Table 3.11: Pathways enriched in $CD8^+$ T cells between AS patients and healthy controls using differentially expressed genes FDR <0.01, * FDR <0.05.

NK cells

A large number of pathways were significantly enriched in NK cells in AS pathogenesis, shown in Table 3.12. The 3 differentially expressed genes involved in tRNA aminoacylation are *FARS2*, *WARS2* and *PPA2*. FARS2 and WARS2 are aminoacyl-tRNA synthetases (AARSs), which catalyse the aminoacylation of tRNA molecules, and are essential for decoding genetic information during the process of translation. tRNA are tightly connected to a vast number of signalling pathways, and are believed to have biological functions other than in protein biosynthesis (Park and Kim 2006) with the AARSs being implicated in the aetiology of many diseases (Park et al. 2008).

Pathway	Genes
Mitochondrial tRNA aminoacylation	*FARS2, WARS2, PPA2*
S1P2 pathway	*JUN, MAPK1, GNAI2*
FGF signaling pathway	*JUN, MAPK1, STAT5B, PLAU*
Mechanism of Gene Regulation by Peroxisome Proliferators via PPARα)	*JUN, MAPK1, STAT5B, MED1*
TNFR1 Signaling Pathway	*JUN, ARHGDIB, CRADD*
ATF-2 transcription factor network	*JUN, MAPK1, PLAU, ARG1*
Validated transcriptional targets of AP1 family members Fra1 and Fra2	*JUN, PLAU, NOS3*
Extracellular matrix organization	*CTSG, ELANE, COL17A1, PCOLCE2*
Angiopoietin receptor Tie2-mediated signaling	*MAPK1, STAT5B, NOS3*
Beta1 integrin cell surface interactions	*PLAU, ITGA9, FBN1*
CXCR4-mediated signaling events	*STAT5B, CFL1, ITGA9, GNAI2*
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	*HLA-C, CD8B, CD200R1*
AP-1 transcription factor network	*JUN, PLAU, FOSB*
T cell receptor signaling pathway	*JUN, MAPK1, CD8B, PDCD1*
Integrins in angiogenesis	*MAPK1, PI4KB, COL17A1*
CDC42 signaling events	*JUN, MAPK1, CFL1*
VEGF signaling pathway	*MAPK1, NOS3, HSPB1*
Cell adhesion molecules (CAMs)	*HLA-C, CD8B, ITGA9, PDCD1*

Table 3.12: Pathways enriched in NK cells between AS patients and healthy controls using differentially expressed genes FDR <0.01.

NF-κB signalling is enriched in 3 immune cell types

The canonical NF-κB pathway was found enriched in 3 cell types; $CD14^+$ monocytes, $CD16^+$ neutrophils and $CD19^+$ B cells. The gene expression

differences in AS patients compared to controls are shown in the manually curated KEGG pathways in Figures 3.3, 3.4 and 3.5. This classical pathway is normally triggered in response to microbial infections and pro-inflammatory cytokines such as TNF-α. TNF Alpha Induced Protein 3 (*TNFAIP3*) is a ubiquitin editing enzyme, shown to have potent anti-inflammatory properties mediated by NF-κB inhibition. In all 3 cell types, this gene was significantly down-regulated in AS patients, therefore mediating an inflammatory phenotype. Another gene involved in ubiquitin mediated proteolysis is *UBE2L3*, which ubiquitinates the NF-κB1 precursor for its rapid degradation. *UBE2L3* was found down-regulated in $CD16^+$ neutrophils and $CD19^+$ B cells, potentially reinforcing the important role of genes involved in ubiquitination in the control of NF-κB signalling. *UBE2L3* found at the 22q11 locus is also a potential AS GWAS associated gene (Cortes et al. 2013). These two differentially expressed genes may highlight the importance of the cross-talk between NF-κB signalling and ubiquitin proteolysis pathways.

The differentially expressed miRNAs *miR-146a-5p* and *miR-16-1-3p* in $CD14^+$ monocytes discussed later in this Chapter are known to regulate the NF-κB pathway, providing further reinforcement of the importance of NF-κB regulation in AS and other inflammatory diseases. The locations of these miRNAs in regards to the manually curated KEGG pathway are shown in Figure 3.3. The *miR-146a-5p* also targets *PTGS2*, which is significantly up-regulated in AS patients as expected since it is the target for NSAIDs. A diagrammatic representation of the miRNAs and the genes up and down-regulated in our data is shown in Figure 3.6.

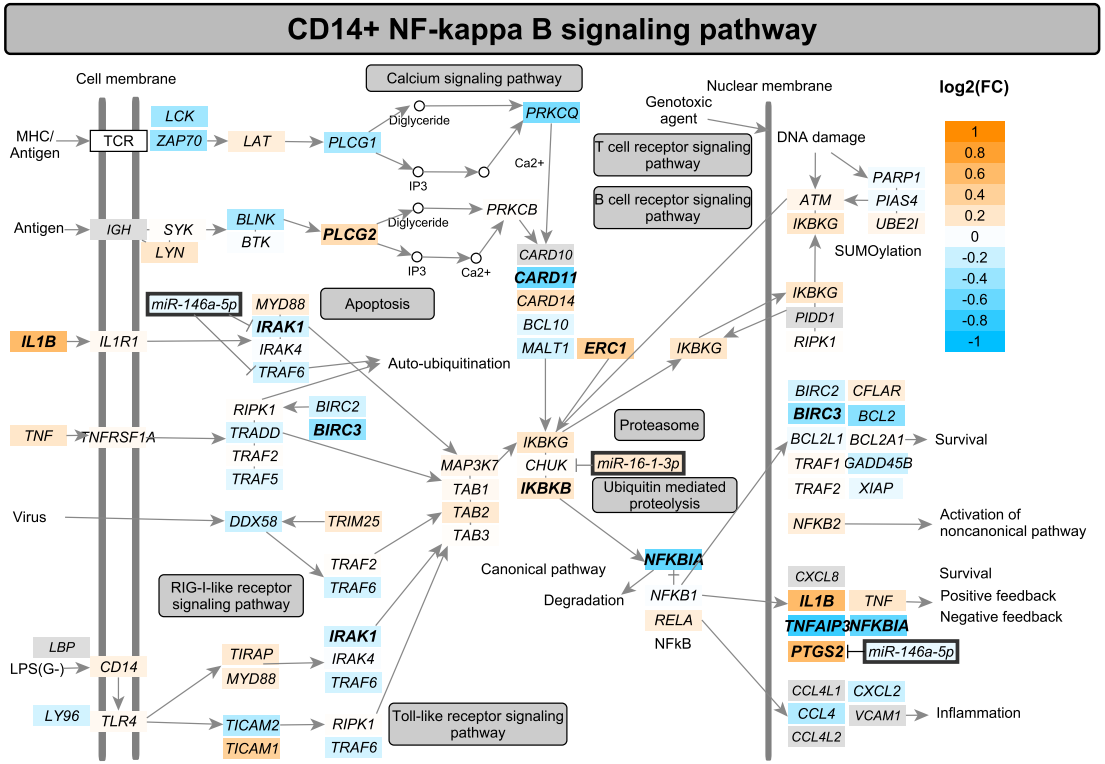


Figure 3.3: NF-κB signalling expression data in AS for CD14$^+$ monocytes using the manually curated KEGG pathway. The genes are coloured by their log2 fold change, demonstrating either up or down-regulation, and the significantly differentially expressed genes are highlighted in bold (FDR <0.05). Shown are also the locations of the differentially expressed miRNAs *miR146a-5p* and *miR-16-1-3p* in NF-κB signalling discussed later in this Chapter, highlighted by the bold outline.

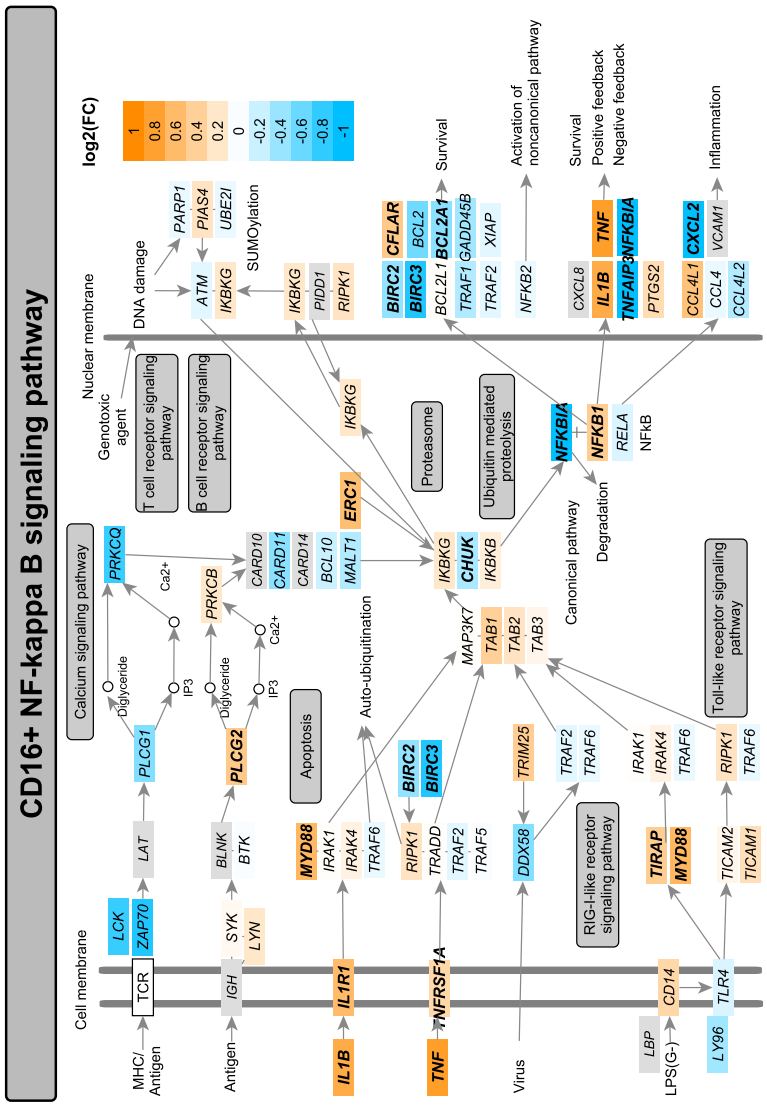


Figure 3.4: NF-κB signalling expression data in AS for CD16$^+$ neutrophils using the manually curated KEGG pathway. The genes are coloured by their log2 fold change, demonstrating either up or down-regulation, and the significantly differentially expressed genes are highlighted in bold (FDR <0.05).

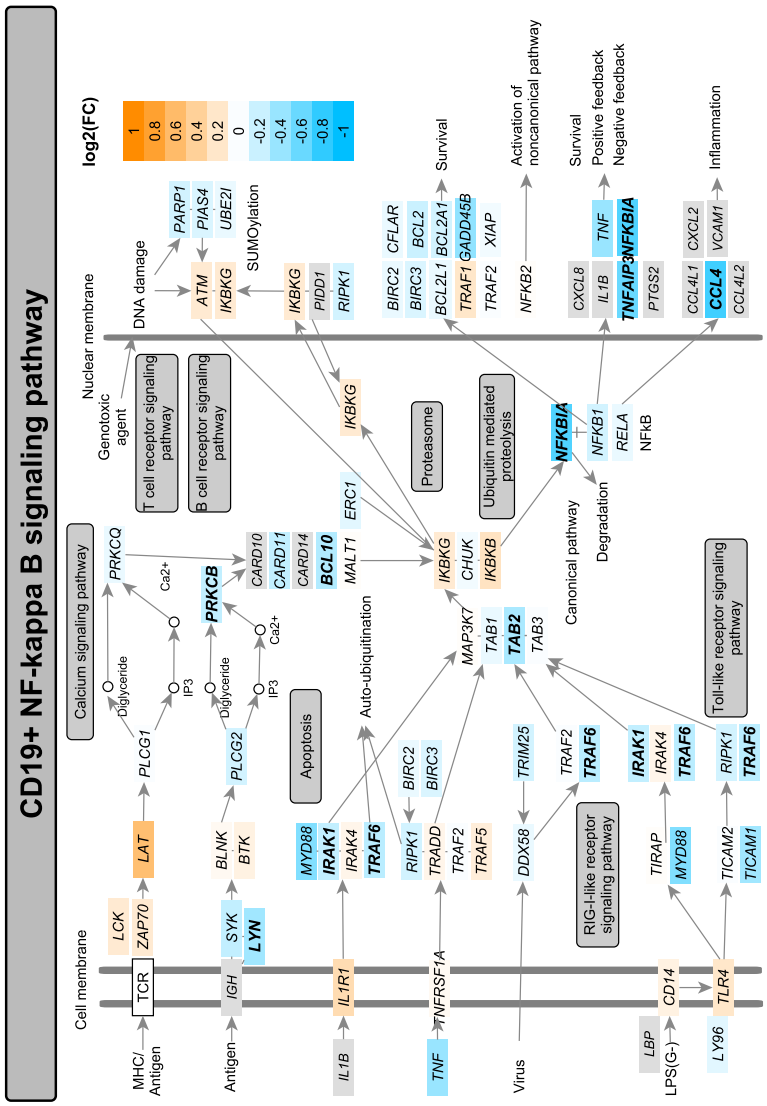


Figure 3.5: NF-κB signalling expression data in AS for CD19$^+$ B cells using the manually curated KEGG pathway. The genes are coloured by their log2 fold change, demonstrating either up or down-regulation, and the significantly differentially expressed genes are highlighted in bold (FDR <0.05).

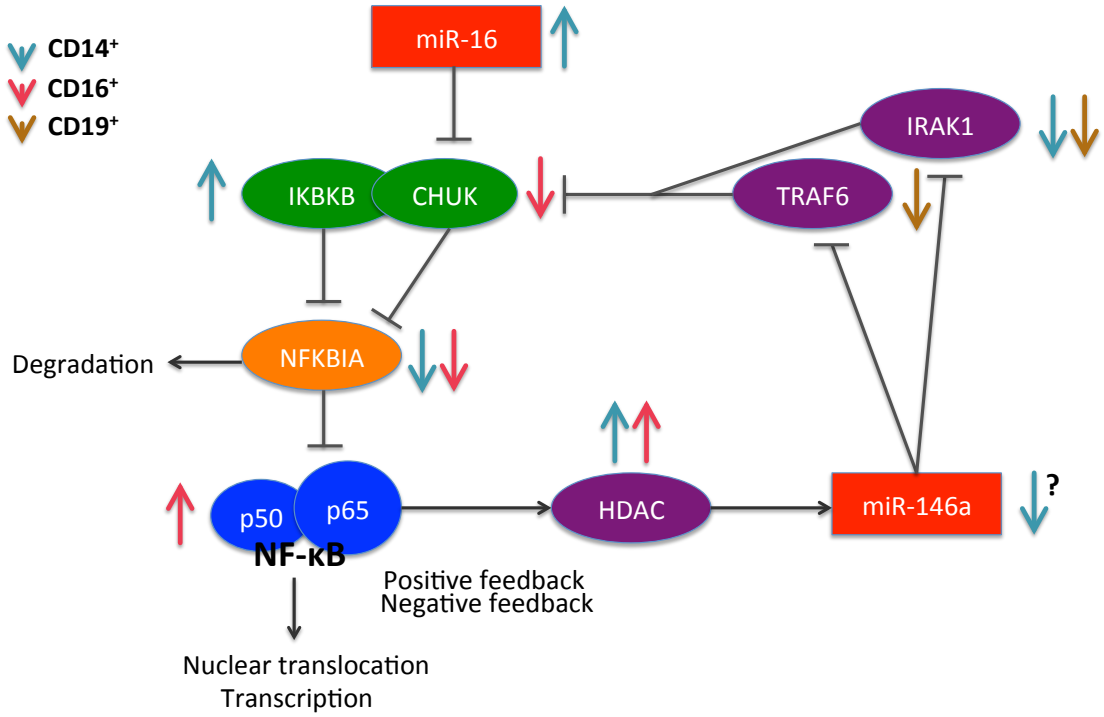


Figure 3.6: Diagrammatic representation of how miRNAs, and specifically *miR-146a* and *miR-16* may be involved in NF-κB signalling, along with the up/down-regulation of each member of the canonical NF-κB signalling pathway in the CD14$^+$, CD16$^+$ and CD19$^+$ data. *MiR-146a* inhibits its targets *TRAF6* and *IRAK1*, which in CD14$^+$ and CD19$^+$ are significantly down-regulated. These increased levels will lead to the inhibition of IKKα or CHUK, which is also the target of *miR-16*, found significantly up-regulated in CD14$^+$ monocytes. *CHUK* works with *IKBKB* to inhibit *NFKBIA*, which in all cell types is significantly down-regulated. This enables NF-κB to translocate to the nucleus and begin transcription of its target genes, such as *IL1B*. *MiR-146a* is positively regulated by NF-κB, meaning that increases of *miR-146a* will eventually work to down-regulate inflammation. Diagram edited and reproduced from (Ma et al. 2011).

HDAC Class I signalling is enriched in innate immune cell types

HDAC Class I signalling was found as an enriched pathway in CD14$^+$ monocytes and CD16$^+$ neutrophils at an FDR <0.01, and in CD8$^+$ T cells with a significance of FDR <0.05. Some of the genes highlighted in this pathway are histone acelytransferases (HATs), which are associated with enhanced rates of gene expression and histone deacetylases (HDACs) which are associated with repressed gene expression. The differentially expressed HATs in both cell types

are *EP300* and *CREBBP*, and the differentially expressed HDACs are *HDAC4*, *HDAC8* and *HDAC10* in $CD14^+$ monocyes, and *HDAC4* and *HDAC7* in $CD16^+$ neutrophils. All of these genes are up-regulated in AS patients versus healthy controls, which is the reverse to a previous study (Toussirot et al. 2013). In both cases, the balance between the two sets of enzymes was the same; the functional relevance of this being unknown. The genes found in the HDAC Class I pathway (FDR <0.05) are shown in Table 3.13.

HDAC Class I signalling	Genes
$CD14^+$	*NFKBIA, EP300, CREBBP, HDAC10, HDAC4, HDAC8* *NCOR2, PRKACA, MAX, RAN, SAP30, SAP18* *GATAD2B, RBBP7, TFCP2*
$CD16^+$	*NFKB1, NFKBIA, STAT3, EP300, CREBBP, HDAC7* *HDAC4, NCOR2, TNF, TNFRSF1A, PRKACA, SUMO1* *SSPO, FKBP3, SAP30, SAP18, GATAD2B, WDR77* *RBBP7, TFCP2, GATAD2A*

Table 3.13: Differentially expressed genes (FDR <0.05) in the HDAC Class I signalling pathway in $CD14^+$ monocytes and $CD16^+$ neutrophils.

Differential lncRNA analysis

$CD16^+$ neutrophils demonstrated the largest number of differentially expressed lncRNAs between AS patients and healthy controls, and $CD8^+$ T cells the lowest. The large majority of lncRNAs do not have functional annotations therefore the interpretation of these results is challenging. Innate immune cells $CD14^+$ monocytes and $CD16^+$ neutrophils demonstrated a significant up-regulation of the nuclear enriched abundant transcript 1 *NEAT1* in patients, consistent with previous studies in SLE and RA (Zhang et al. 2016; Song et al. 2015). The growth arrest-specific transcript 5 *GAS5* was found down-regulated in both cell types, and is found within a known GWAS locus for SLE (Suarez-Gestal et al. 2009).

SNHG5 was found differentially expressed in all cell types apart from NK T cells. A number of lncRNAs relating to the spliceosome are found differentially expressed in NK T cells, including *U1*, *RNU11* and *RNU12*. A list of currently well annotated lnRNAs found differentially expressed (FDR <0.05) are listed in Table 3.14.

CD14$^+$	**CD16$^+$**	**CD19$^+$**	**CD4$^+$**	**CD8$^+$**	**NK**
SNHG5	*SNHG1*	*XIST*	*XIST*	*XIST*	*U1*
FTX	*DLEU2*	*SNHG14*	*TSIX*	*SNHG5*	*RNU11*
HCG18	*SNHG15*	*DLEU2*	*SNHG5*	*KMT2B*	*RNU12*
DLEU2	*NEAT1*	*HMGA1P4*	*RNU12*		*HCG14*
KMT2B	*SNHG5*	*PWAR6*			*FTX*
GAS5	*SNHG8*	*UHRF1*			
XIST	*BCYRN1*	*SNHG5*			
GHRLOS	*GAS5*	*SCARNA9*			
CASC7	*PSORS1C3*	*MALAT1*			
SCARNA2	*FTX*	*RNU12*			
LUCAT1	*KCNQ1OT1*	*KCP*			
ZFAS1	*CACS7*	*FAM66C*			
TSIX	*7SK*				
JPX	*SEMA3B*				
FAM66C	*TERC*				
RMRP	*XIST*				
NEAT1	*KMT2B*				
	SNHG11				

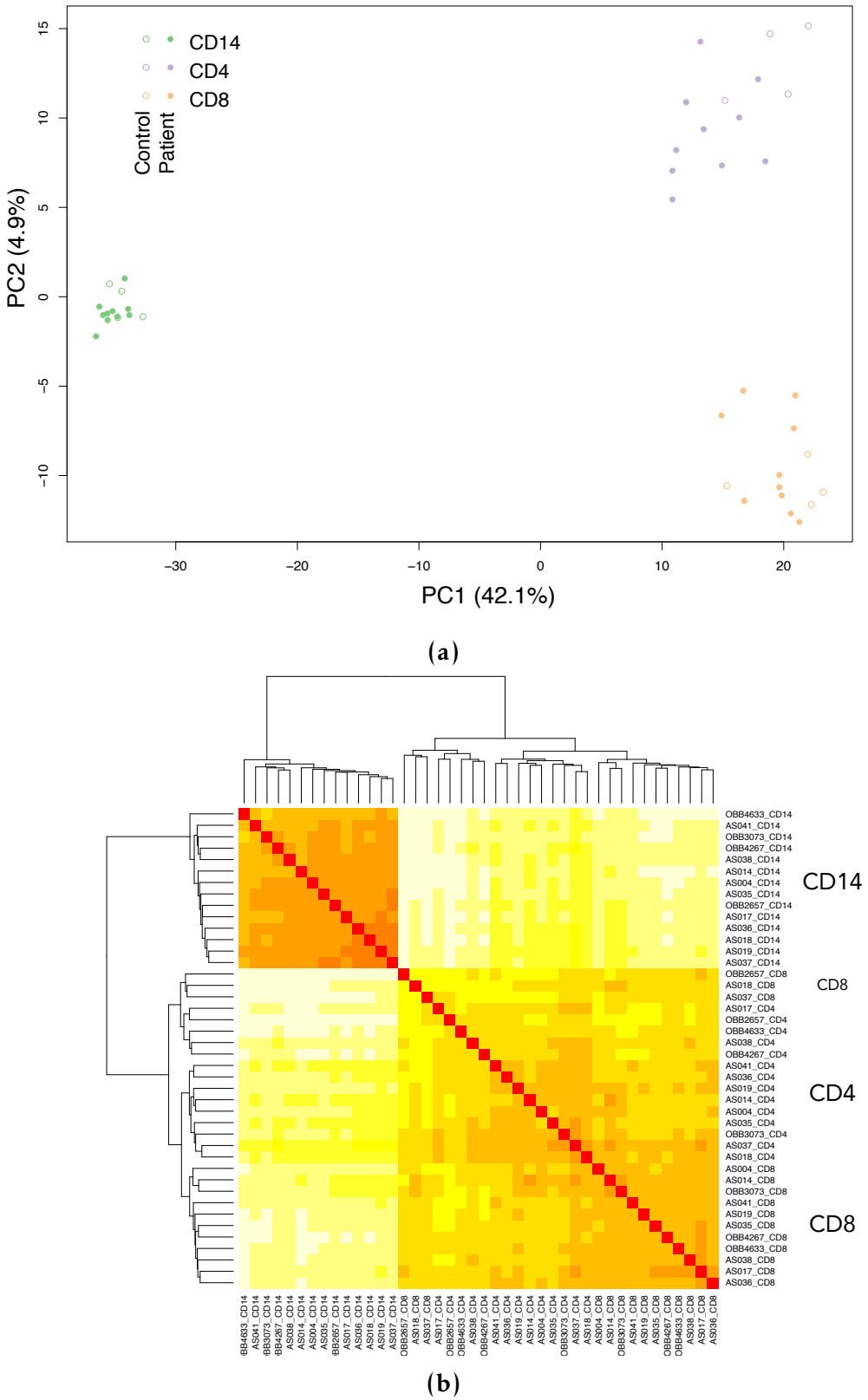
Table 3.14: Differentially expressed lncRNAs with functional annotations in each cell type. (FDR <0.05).

3.3.3 Small RNA-seq (miRNA) differential expression in AS

Small RNA-seq was carried out for the four healthy controls from the first round of recruitment, and the 10 AS patients (Tables 3.2 and 3.3).

Quality control

The nature of the protocol for isolating miRNAs may lead to variable mapping rates to known miRNAs due to the size selection protocol, meaning that the complexity of the small RNA library may be low due to the limited number of unique small RNA sequences. The number of reads mapping to miRNAs in this sample set varied between 1.5 million to 7 million. This is not unusual for miRNA projects, and the Oxford Genomics Centre at the Wellcome Trust Centre for Human Genetics have also reported a great variability in mapping rates, varying from over 70% down to 20%. Plotting the first two principle components of variance separates the cell types to form distinct clusters (Figure 3.7(a)), and unsupervised hierarchical clustering demonstrates that $CD14^+$ monocytes form their own distinct cluster, while $CD4^+$ and $CD8^+$ T cells are more similar but still mostly distinguishable (Figure 3.7(b)). Due to the small sample size and no obvious outliers, all samples were included for downstream analysis.



(a)

(b)

Figure 3.7: Quality Control of miRNA samples. **(a)** Principle component analysis plotting the first two PCs. **(b)** Distance matrix heatmap using unsupervised hierarchical clustering. The main variance is between monocytes and T cells, with a clear separation between cell types.

Differential miRNA expression analysis

Differential expression analysis was carried out between AS patients and controls for the 3 cell types using DESeq2 (Love et al. 2014), using all miRNA transcripts with counts >1. The multi-factor design function within DESeq2 was used to account for age, sex and disease activity based on the BASDAI score as a linear covariate. The results are shown in Table 3.15. Only one differentially expressed miRNA between AS patients and controls was found for $CD14^+$ monocytes and $CD8^+$ T cells, and 5 for $CD4^+$ T cells. Only one miRNA was differentially expressed in two cell types. *MiR-100-5p* was found as the top differentially expressed miRNA in both T cell subtypes, at high significance. The most striking differentially expressed miRNA is *miR-146a-5p*, which is already reported to have a role in AS and related diseases (Chatzikyriakidou et al. 2010; Qian et al. 2016; Xu et al. 2015; Bogunia-Kubik et al. 2016; Chatzikyriakidou et al. 2010; Zhang et al. 2017), and has an important role in inflammation (Lawrence et al. 2005).

Taking into account disease activity based on the BASDAI score may introduce some bias and variability between individuals, since it is a self reported measure of severity and pain. Therefore the analysis may also be as informative when removing this covariate.

$CD14^+$	FDR	$CD4^+$	FDR	$CD8^+$	FDR
miR-146a-5p	0.02	miR-100-5p	$1.70x10^{-08}$	miR-100-5p	$6.45x10^{-23}$
				miR-3653	$1.71x10^{-06}$
				miR-26a-2-3p	0.005
				miR-126-3p	0.02
				miR-99a-5p	0.02

Table 3.15: miRNAs differentially expressed in $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells between AS patients and healthy controls accounting for disease activity. The analysis was carried out accounting for age, sex and disease activity as a linear covariate. These miRNAs may therefore act as a disease activity marker.

$CD14^+$	FDR	$CD4^+$	FDR	$CD8^+$	FDR
miR-16-1-3p	0.04	miR-4443	$2x10^{-05}$	-	-
miR-148a-3p	0.05	miR-342-5p	0.03		

Table 3.16: miRNAs differentially expressed in $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells between AS patients and healthy controls without accounting for disease activity.

$CD14^+$	FDR	$CD4^+$	FDR	$CD8^+$	FDR
-	-	miR-543	0.025	-	-
		miR-582-3p	0.025		
		miR-370-3p	0.036		
		miR-409-3p	0.036		

Table 3.17: miRNAs differentially expressed between patients with high or low disease activity. Patients GNORLAS004, 14, 17, 18, 19 and 35 compared with GNORLAS036, 37, 38 and 41.

$CD14^+$	FDR	$CD4^+$	FDR	$CD8^+$	FDR
miR-708-3p	0.01	miR-4443	0.0001	miR-3615	0.043
miR-16-1-3p	0.067	miR-409-3p	0.0006		
miR-4443	0.17	miR-144-3p	0.014		
		miR-451a	0.03		
		miR-5100	0.043		
		miR-224-5p	0.057		

Table 3.18: miRNAs differentially expressed between patients with high disease activity and healthy controls. Patients GNORLAS004, 14, 17, 18, 19 and 35 compared with healthy controls.

$CD14^+$	FDR	$CD4^+$	FDR	$CD8^+$	FDR
miR-146a-5p	$5.95x10^{-05}$	miR-4443	$1.21x10^{-05}$	-	-
miR-16-1-3p	0.013	miR-342-5p	0.00014		
		miR-143-3p	0.005		
		miR-30a-5p	0.008		
		miR-2110	0.009		
		miR-30e-5p	0.017		

Table 3.19: miRNAs differentially expressed between patients with low disease activity and healthy controls Patients GNORLAS036, 37, 38 and 41 compared with healthy controls.

Repeating the analysis and removing the disease activity covariate resulted in a change of differentially expressed miRNAs (Table 3.16), with *miR-146a-5p* losing its significance (FDR = 0.77). *MiRNA-146a-5p* may therefore be correlated with disease activity, and plotting the normalised counts for *miR-146a-5p* demonstrate a trend in expression, with the 4 healthy controls exhibiting the highest expression, and all AS patients exhibiting a variety of lower normalised counts (Figure 3.8). *MiR-146a-5p* is downregulated in AS patients, similar to SLE patients (Tang et al. 2009; Lofgren et al. 2012; Ji et al. 2014). The greatest difference in expression exists between controls and patients with low disease activity, and when differential expression is tested comparing AS low disease activity patients versus healthy controls, *miR-146a-5p* increases in significance 5.95 x 10^{-05} (Table 3.19). The disease activity marker miR-100-5p also loses significance when removing the disease activity covariate.

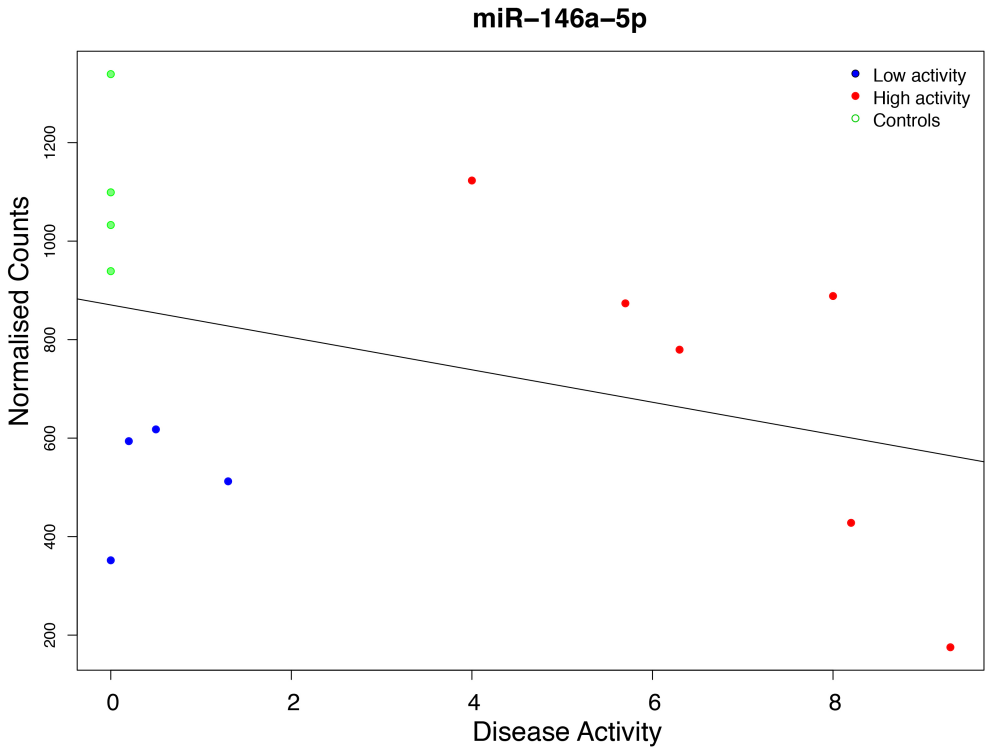


Figure 3.8: CD14$^+$ *miR-146a-5p* normalised counts in AS patients versus controls. The trend line demonstrates a down-regulation of *miR-146a-5p* in AS in CD14$^+$ monocytes, with the largest difference in the low disease activity group. However this correlation is not significant (r^2= 0.13, P=0.21).

Removing disease activity from the design formula gives us an additional differentially expressed miRNA *miR-16-1-3p* in CD14$^+$ monocytes (Table 3.16). *MiR-16* has previously been found to be up-regulated in AS T-cells (Lai et al. 2013), instead of being up-regulated in CD14$^+$ monocytes as it is here (Figure 3.9. Both *miRNA-146a-5p* and *miR-16* are involved in the regulation of NF-κB signaling (Ma et al. 2011), and could therefore reinforce NF-κB signalling as a vital pathway in AS pathogenesis, as previously shown in Figure 3.3.

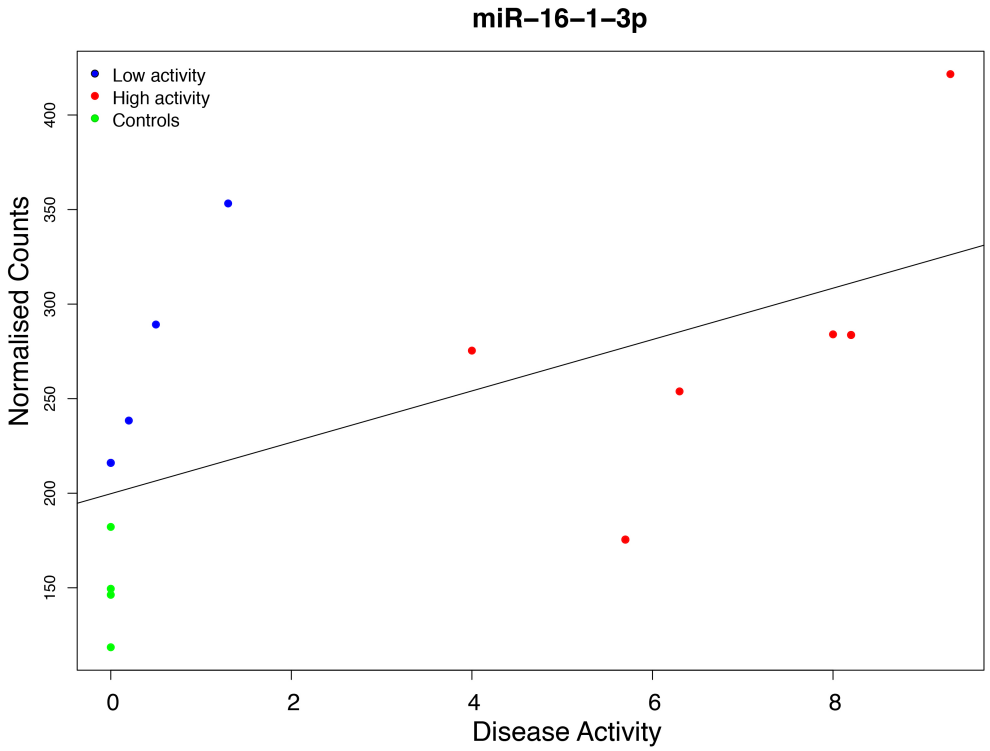


Figure 3.9: CD14$^+$ *miR-16-1-3p* normalised counts in AS patients versus controls. The trend line demonstrates an up-regulation of *miR-16-1-3p* in AS in CD14$^+$ monocytes, with the highest counts in the high activity AS patient group (r^2= 0.34, *P*=0.028).

Repeating the analysis with subsets of AS patients separated by disease activity against healthy controls reveals *miRNA-146a-5p* and *miR-16-1-3p* as differentially expressed in CD14$^+$ monocytes (Table 3.18). Two previously reported miRNAs are also found differentially expressed in CD4$^+$ T cells; *miR-409-3p* and *miR-30e-5p* (Lai et al. 2013). Comparing AS patients with low and high disease activity found no significant results for CD14$^+$ monocytes but CD4$^+$ T cells demonstrated a difference in *miR-409-3p* once again (Table 3.17). *MiR-409-3p* was previously found to be down-regulated in AS T cells (Lai et al. 2013), but in this case is up-regulated in the high disease activity AS group in both analyses (Figure 3.10).

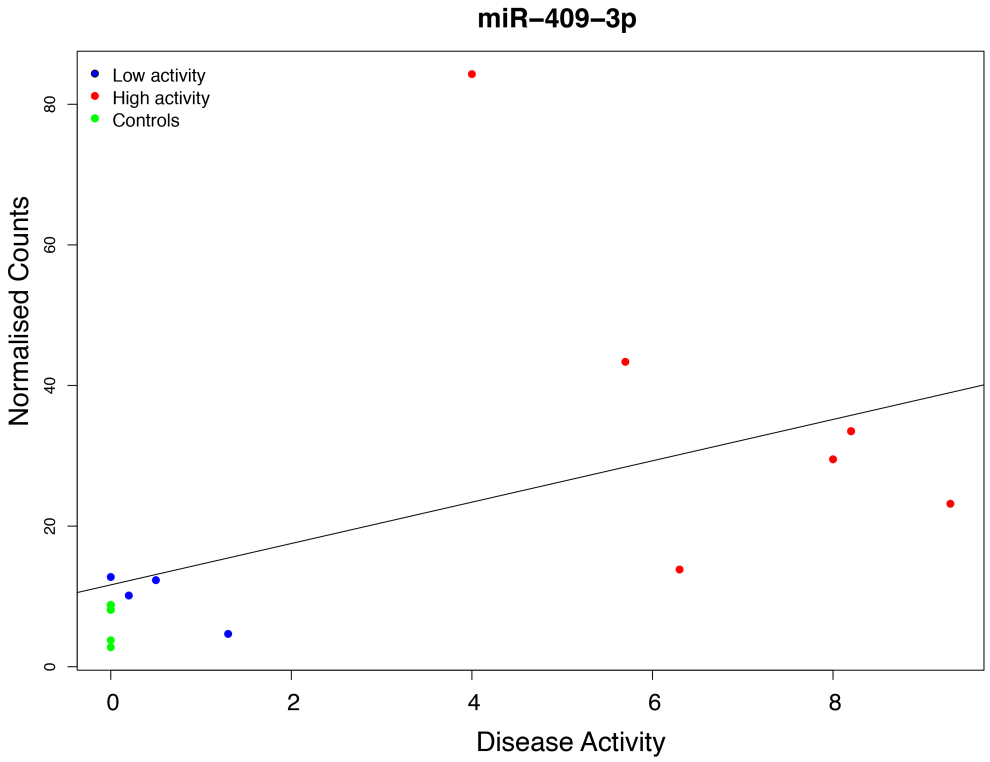


Figure 3.10: CD4$^+$ T cells *miR-409-3p* normalised counts in AS patients versus controls. The trend line demonstrates an up-regulation of *miR-409-3p* in AS in CD4$^+$ T cells, with the highest expression in the high disease activity AS patients (r^2= 0.24, P=0.076).

There is a visible trend in the expression of these 3 miRNAs between AS patients and controls, and even between AS patients with high and low disease activity. The addition of further AS patients and healthy controls to this analysis in the future will provide us with more confidence in these differentially expressed miRNAs, as in some cases they do not reach significance.

miRNA gene targets

Understanding miRNA function has focused on the recognition of their target mRNAs, with the aim to discover which genes they may be down-regulating. The ability of a miRNA to translationally repress a target mRNA is largely dictated by its ability to bind to the first eight nucleotides, which is often referred to as seed nucleotides (Doench and Sharp 2004). There are a number of computational prediction sites that can be used to identify gene targets such as targetscan

(Grimson et al. 2007), (http://www.targetscan.org), as well as databases that report validated gene targets from experimental validation and NGS. I decided to only include mRNA targets with strong evidence in miRTarBase (Chou et al. 2016), (http://mirtarbase.mbc.nctu.edu.tw/index.php). The number of target genes are not enough to be able to carry out pathway analysis, but examining experimentally validated genes can begin to give us an indication of important genes and pathways involved in AS. The largest list of experimentally validated mRNA targets is for *miR-146a-5p*, due to its known role in the immune system. The genes listed again suggest the NF-κB signalling pathway, and other immune system genes such as *CXCR4* (Table 3.20).

Table 3.20: mRNA target with strong evidence from miRTarBase.

miRNA	mRNA targets
miR-146a-5p	*CXCR4, CFH, CFH, IRAK2, TLR2, FADD, TRAF6x15, IRAK1x15, ROCK1, BRCA2, BRCA1x3 FAF1, CCNA2, PA2G4, IL8, NFKB1x8, CDKN1A, EGFRx6, MTA2,CD40LG, FASx2, ERBB4x2 SMAD4x5, TLR4, WASF2, STAT1, UHRF1, SPP1, SLPI, L1CAM, SMN1, CARD10x2, COPS8 ELAVL1, NUMBx2, ICAM1, SMAD2, PTGS2, CCL5, ROBO1, SIKE1, CXCL12, PRKCE, HOXD10 RAC1, LAMC2, COX2, DUSP1, RNF11*
miR-100-5p	*PLK1, EGR2, ID1, FGFR3, MMP13, ATM, IGFR1, BMPR2, CTDSPL, FLT1, MTOR*
miR-16-1-3p	*CCND1, CCNE1, NFKB1, IKKα*
miR-126-3p	*CRK, VEGFA, SPRED1, PIK3R2, IRS1, PLK2, TOM1, SLC7A5, ADAM9, CXCR4, EGFL7 RGS3, SOX2, DNMT1, PIK3CG, CRKL, SLC41A2, PITPNC1, IGFBP2, KRAS, SLC45A3, CCNE2 HOXA9, MERTK, VCAM1, E2F1, TWF1, TWF2, PTPN7, TEK, MMP7, CXCL12, PGR*
miR-99a-5p	*FOXO3, BCL2, SPRED1, Camsap1, CD97, TCF4, Cdkn1b, RHOU, LRP6, ADM, SIRT1, NFKBIA RAVER2, FGFR3, SERPINE1, IGF1R, MTOR, AGO2*
miR-148a-3p	*DNMT1, HLA-G, TGIF2, DNMT3B, Camk2a, NR1I2, RPS6KA5, Kdm6b, CCKBR, Rock1, IRS1 ACVR1, BCL2, TMED7, Dnmt1*
miR-4443	only ”weak” evidence (NGS/microarray)

miR-342-5p	only ”weak” evidence (NGS/microarray)
miR-3653	only ”weak” evidence (NGS/microarray)
miR-26a-2-3p	only ”weak” evidence (NGS/microarray)
AS patients subset analysis	
miR-543	*PTK2, TWIST1, TWIST1*
miR-582-3p	only weak predictions in miRBase
miR-370-3p	*MAP3K8, HMGA2, CPT1A, TGFBR2, NF1, FOXO1, FOXO1, FOXM1, FOXM1, FOXM1*
miR-409-3p	*FGB, FGA, FGG, PHF10, ANG, RDX, IFNG, MET*

Evidence for the roles of differentially expressed miRNAs using mRNA expression data

The integration of miRNAs and their target mRNAs can be informative in understanding possible transcriptional regulation mechanisms in disease, and how they may differ across cell types. The miRNA analysis may lack power due to its small sample size, but by intersecting the experimentally validated target genes with differentially expressed genes, a small number of targets were identified that may be regulated by these miRNAs in AS. These mRNA targets are shown in Table 3.21, with the majority of results in $CD14^{+}$ monocytes for *miR-146a-5p*. *IRAK1* is a direct target of *miR-146a-5p*, and is found down-regulated in AS patients versus controls. *NFKBIA* is found downstream of *IRAK1* and is also found significantly down-regulated in AS patients. These miRNAs in the context of the canonical NF-κB pathway are shown in Figure 3.3. *COPS8* forms part of a complex also involved in the phosphorylation of *IκBα* and *NFKBIA*.

Further miRNA targets that are differentially expressed include *PA2G4*. The gene product is a transcriptional co-repressor of cell cycle regulatory genes through its interactions with histone deacetylases. WASF1 forms part of a multi-protein complex that can transduce signals to change the shape and mobility of monocytes. CCL5 is a chemoattractant for monocytes, and CXCR4 a chemokine involved in the induction of MAPK signalling known to show aberrant expression in patients with juvenile spondyloarthritis (Lamot et al. 2014). PTGS2, also known as COX-2, is responsible for production of inflammatory prostaglandins. *PTGS2* is found up-regulated in AS patients, which is associated with increased cell adhesion, phenotypic changes and resistance to apoptosis. PTGS2 is an approved drug target for ankylosing spondylitis, known as NSAIDs, that act by binding to the cyclooxygenase (COX) active site.

The one target mRNA for *miR-148a-3p*, found differentially expressed without correcting for disease activity in monocytes was *KDM6B*; its gene product is involved in the inflammatory response by participating in macrophage differentiation in response to inflammation (De Santa et al. 2007). This analysis is limited to the number of experimentally validated miRNA targets, and using predicted mRNA targets instead would give a larger amount of genes, but may increase false positives as miRNAs are predicted to bind to hundreds of genes.

mRNA	log2FC	FDR
CD14$^+$		
miR-146a-5p		
NFKBIA	-0.60	0.0007
PA2G4	-0.24	0.006
WASF2	0.33	0.008
CCL5	-0.68	0.015
COPS8	-0.26	0.019
CXCR4	-0.50	0.029
PTGS2	0.59	0.034
IRAK1	-0.16	0.043
SMN1	-0.35	0.055*
miR-148a-3p		
KDM6B	0.64	0.007
CD8$^+$		
miR-126-3p		
NFKBIA	-0.64	0.008

Table 3.21: Integration of differentially expressed mRNAs targeted by the differentially expressed miRNAs. Overlap between miRNA targets and differentially expressed mRNA was only demonstrated for 3 miRNAs, *miR-146a-5p*, *miR-148a-3p* and *miR-126-3p*. *One target just exceeded the FDR of <0.05.

3.3.4 Differential expression analysis by HLA-B*27 genotype in healthy controls

The *HLA-B*27* genotype is known to be strongly associated with AS, with the majority of AS patients having at least one B*27 allele. The effect of B*27 on global gene expression was tested in healthy individuals, comparing *HLA-B*2705* heterozygotes against *HLA-B*0801* homozygote controls (Table 3.3). The differential expression analysis was carried out using DESeq2, correcting for sex. The number of differentially expressed genes across cell types are limited, with $CD4^+$ T cells demonstrating the largest number (Table 3.22).

In $CD14^+$ monocytes *NR4A1* was found to be down-regulated in B*27 individuals. This nuclear receptor has been implicated as a negative regulator of NF-κB signalling, and is known to promote anti-inflammatory effects in macrophage apoptotic cells and is necessary in the maintenance of self-tolerance in the murine model of lupus (Ipseiz et al. 2014). *NOV*, also known as *CCN3* was found differentially up-regulated in $CD16^+$ neutrophils in B*27 homozygotes. Its gene expression is regulated by TNF-α and IL1 cytokines, and has a potential role in inflammatory disease (Perbal 2006).

$CD4^+$ and $CD8^+$ T cells both demonstrated down-regulation of *CDKN1A* in B*27 homozygotes compared with B*0801 individuals. CDKN1A functions as a regulator of cell cycle progression and the protein is reported to be cleaved by CASP3-like caspases, and therefore involved in apoptosis signalling (Woo et al. 2003). The $CD4^+$ comparison also shows a down-regulation of *NFKBIE*, which is an inhibitor of NF-κB signaling. The comparison in $CD19^+$ B cells found a down-regulation of *CDKNIC*, a negative regulator of proliferation, and also the down-regulation of *SOC3*. The gene product of *SOC3* is a suppressor of cytokine signalling and inhibits the activity of JAK2 kinase during JAK/STAT signalling (Babon et al. 2012). The B*27 homozygotes in NK cells demonstrated a down-

regulation of *BCL2*, which is an apoptosis regulator. The differentially expressed genes for all cell types are demonstrated in Figure 3.11. A top down-regulated gene in $CD16^+$ neutrophils, $CD4^+$ and $CD8^+$ T cells was *XXBAC-BPG181B23.7*, also known as *lnc-HLA-B-2*. No functional information is listed for this lncRNA. Pathways analysis was carried out in $CD4^+$ T cells using XGR, and enriched pathways are shown in Table 3.23.

Cell Type	B*2705 homozygotes
	FDR 0.05/0.01
$CD14^+$	21/15
$CD4^+$	76/46
$CD8^+$	15/9
$CD16^+$	14/8
$CD19^+$	23/16
NK	20/19

Table 3.22: The number of differentially expressed genes between healthy individuals that are *HLA-B*27* homozygous (n= 5) versus *HLA-B*0801* homozygotes (n= 5). An FDR of <0.05 and 0.01 was used to define differentially expressed genes.

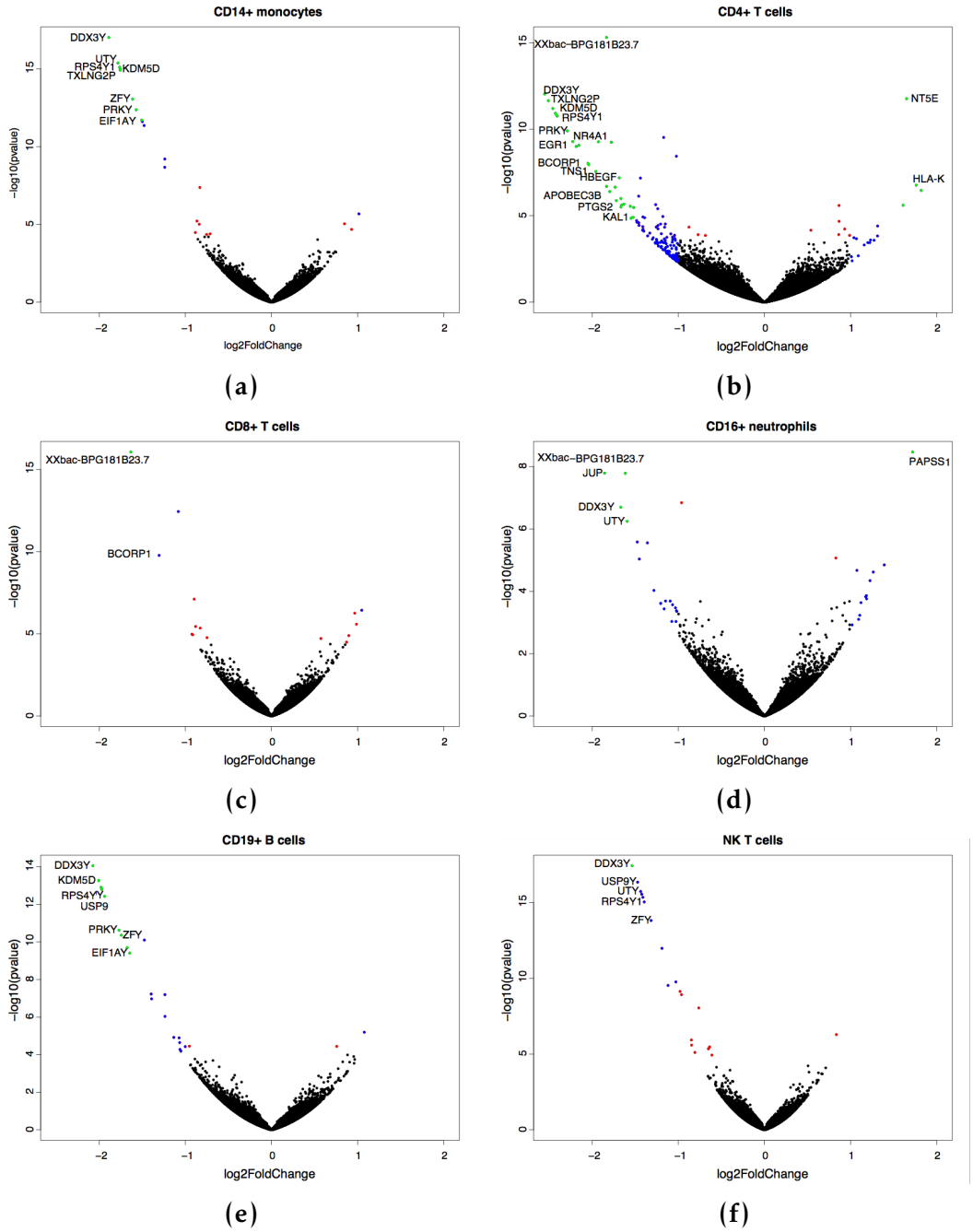


Figure 3.11: Volcano plots of differentially expressed genes in *HLA-B*2705* homozygotes versus *HLA-B*0801* homozygotes in healthy controls. The red points demonstrate genes differentially expressed with an FDR <0.05, blue points with a Log2FoldChange >1 and green points with a Log2FoldChange >1.5 in **(a)** $CD14^+$ monocytes **(b)** $CD4^+$ T cells **(c)** $CD8^+$ T cells **(d)** $CD16^+$ neutrophils **(e)** $CD19^+$ B cells **(f)** NK T cells.

Table 3.23: Pathways enriched in CD4$^+$ T cells in healthy individuals with one or more copies of *HLA-B*2705* compared with *HLA-B*0801*.FDR <0.01.

B*2705 homozygotes	Genes
AP-1 transcription factor network	*FOS, EGR1, TCF7L2, ATF3*
Calcineurin-regulated NFAT-dependent transcription in lymphocytes	*FOS, EGR1, PTGS2*
Interferon alpha/beta signaling	*EGR1, HLA-C, HLA-K*
Glucocorticoid receptor regulatory network	*FOS, EGR1, NR4A1*
Immune System	*FOS, EGR1, NFKBIE, HLA-C, NR4A1, HLA-K*
Pathways in cancer	*FOS, PTGS2, TCF7L2*

3.4 Discussion

This results chapter presented the identification of differentially expressed genes between healthy controls and AS patients in six cell types. Pathway enrichment analysis has highlighted the importance of NF-κB signalling, HDAC Class I signalling and T cell receptor signalling among others. The potential importance of lncRNAs in AS pathogenesis was also highlighted, with known autoimmune disease lncRNAs *NEAT1* and *GAS5* demonstrating differential expression. The identification of miRNAs associated with AS disease aetiology, specifically in NF-κB signalling was also described. Finally, the limited differences in gene expression between healthy individuals with different HLA-B*2705 genotypes were highlighted, but demonstrated enrichment in genes relating to NF-κB signalling, cell proliferation and apoptosis.

3.4.1 AS cell type specific gene expression

Analysis of expression datasets combining the analysis of miRNAs, mRNAs and lncRNAs in the same or similar individuals across multiple cell types is a powerful tool in beginning to understand the regulation of AS differentially expressed genes and their relevance in different cell subsets. The complexity of common immune mediated diseases such as AS is highlighted in the genes and enriched pathways that have been found in this and previous studies (Fang et al. 2015; Chen et al. 2013; Assassi et al. 2011; Pimentel-Santos et al. 2011; Duan et al. 2010). The work presented in this chapter has identified various molecular mechanisms that are both shared but also specific to immune cell subsets. While of limited sample size, this is the first study that investigates the gene expression profiles in AS patients in up to 6 cell types, highlighting the cell specific transcriptional landscape relevant to innate and adaptive functions of the immune system in AS aetiology.

Analysis of the differentially expressed genes between AS patients and healthy controls in six cell types has further highlighted the importance of cell specificity in understanding disease mechanisms. Each cell type demonstrated uniquely enriched pathways, but also shared potential mechanisms. These often remained grouped by innate and adaptive responses, such as NF-κB signalling and HDAC Class I signalling in the innate immune cells ($CD14^+$ monocytes and $CD16^+$ neutrophils), and T cell receptor signalling in T lymphocytes belonging to the adaptive arm.

Antigen processing and presentation, an important pathway hypothesised through GWAS (Evans et al. 2011) was only differentially regulated in AS $CD14^+$ monocytes. The downstream effects of antigen presentation may result in NK-mediated cytotoxicity and aberrant T cell receptor signalling. *ERAP1* was not found significantly differentially expressed in monocytes, suggesting that its pathogenic role may be related more to changes in protein structure. Endoplasmic reticulum (ER) genes *PDIA3* and *CALR* were found significantly down-regulated in patients, along with the proteosome subunit gene *PSME1*. PDIA3, also known as ERp57 may be involved in mediating protein folding, and also acts as a molecular chaperone to prevent the formation of protein aggregates. It was shown that the absence of PDIA3 may lead to stable aggregates within the ER, but was not shown to induce the UPR (Stepensky et al. 2007). Its down-regulation may suggest a build up of protein aggregates in the ER. PSME1 is required for immunoproteasome assembly, of which an essential function is the processing of class I MHC peptides. *PSME2* is also down-regulated, although not significantly, but together these proteins may reduce antigen processing that is required by the immunoproteasome.

The ubiquitin proteosome pathway was shown to be important in a proteomics study carried out in AS monocytes (Wright et al. 2009). Only one of the proteins reported to be differentially expressed in the ubiquitin proteosome pathway

overlapped with a differentially expressed gene in AS monocytes, *USP31*. A tumour necrosis factor TNF-α inducible protein PTX3, was also found down-regulated in the RNA-seq data.

IL-8 and CXCR2-mediated signalling is a neutrophil specific enriched pathway that may be relevant in AS pathogenesis. The CXCR2 macromolecular signalling complex, which contains CXCR2, NHERF1 and PLCB2 has been shown to regulate neutrophil functions in inflammatory diseases, and the disruption of this complex was demonstrated to inhibit neutrophil trans-epithelial migration (Wu et al. 2012). *PLCB2* was found up-regulated in AS patients in our data and may be mediating neutrophil migration by coupling with CXCR2.

The mTOR signalling pathway was the top enriched pathway in $CD19^+$ B cells. Class I PI3K signalling events mediated by Akt is a related pathway that was also found significantly enriched, and is a crucial pathway for many aspects of B cell biological processes (Jellusova and Rickert 2016), including B cell receptor (BCR) signalling. BCR signalling and PI3K signalling can activate ERK signalling, important for B cell antigen induced proliferation and development (Jacob et al. 2002).

$CD8^+$ T cells demonstrated an up-regulation of *AGER*. Non-coding variants have been associated with *AGER* expression, with protein levels (RAGE) and polymorphisms associated with CD (Dabritz et al. 2011), RA (Hofmann et al. 2002) and UC (Wang et al. 2016).

Pathway enrichment analysis in NK T cells identified the Sphingosine-1-phosphate (S1P) pathway. This is involved in cell proliferation and inflammation where S1P acts either directly on intracellular targets or activates G protein-coupled receptors. *GNAI2* was found differentially expressed in NK T cells. Its gene product is involved in G protein signal transduction. This gene is also

involved in CXCR4-mediated signalling events, also an enriched pathway in NK T cells.

Very few changes were identified across healthy individuals between *HLA-B*2705* homozygotes compared with *HLA-B*0801* homozygotes, and may demonstrate the limited effect of B*27 in individuals not predisposed to an inflammatory state. However, a few genes relating to autophagy, NF-κB signalling and cell proliferation may be important to investigate, along with the glucocorticoid receptor regulatory network which was an enriched pathway in $CD4^+$ T cells and already targeted by therapies such as corticosteroids.

3.4.2 Evidence for altered NF-κB signalling in AS cell types

The enrichment of NF-κB signalling in $CD14^+$ monocytes, $CD16^+$ neutrophils and $CD19^+$ B cells may demonstrate its importance in AS pathogenesis. SNPs in *NFKB1* were found in a GWAS for IBD (Jostins et al. 2012), and loci encompassing the *NFKB1* genes have shown the most inflammatory disease associations so far (Parkes et al. 2013), suggesting that its regulation is important in the development of inflammatory diseases. All 3 cell types demonstrated a significant down-regulation of *NFKBIA*, an inhibitor of NF-κB, and an up-regulation of *IL-1B*, *TNF*, *PTGS2* and the CASP8 FADD Like Apoptosis Regulator (*CFLAR*). Disease associated SNPs, especially those associated with inflammatory diseases show enrichment in sequences bound by NF-κB (Karczewski et al. 2013). *TNFAIP3* was down-regulated in AS patients in all 3 cell types, suggesting its importance in immune and inflammatory responses signalled by cytokines such as TNF-α and IL-1B, extending the results in an AS microarray study in PBMCs (Duan et al. 2010). *TNFAIP3* codes for the ubiquitin-modifying enzyme A20 which is pleiotropically expressed, and regulated at both the transcriptional and the post-transcriptional level. Animals deficient of *TNFAIP3* in myeloid lineages develop an aggressive form of arthritis (Matmati et al. 2011), and deletion of

the gene in dendritic cells has been shown to cause colitis (Hammer et al. 2011). A20-deficiency in B cells also caused mice to develop an autoimmune condition similar to SLE (Hovelmeyer et al. 2011; Chu et al. 2011), possibly due to exaggerated NF-κB mediated responses resulting in increased germinal centre B cells. *TNFAIP3* has already been associated with other autoimmune diseases in GWAS, such as: SLE, RA, Type 1 Diabetes, CD, coeliac disease, primary biliary cirrhosis, systemic sclerosis and psoriasis (Trynka et al. 2009; Ma and Malynn 2012; Jostins et al. 2012; Cordell et al. 2015). A haplotype associated with SLE risk identified by fine-mapping is located at a potential enhancer site, and demonstrated reduced binding of NF-κB subunits, resulting in reduced expression of *TNFAIP3* (Adrianto et al. 2011). A recent study looking at *TNFAIP3* levels in axial spondyloarthritis (AxSpA) macrophages also found its levels reduced in patients (Liu et al. 2017), correlating with increased LPS-induced TNF-α.

Understanding the association of NF-κB activity and inflammatory disease is not easy to interpret due to the balance between pro- and anti-inflammatory mediators by NF-κB feedback loops. Genetic studies in mice show that NF-κB activation is not necessarily pro-inflammatory and has complex roles in the inflammatory response, and is most likely playing a tissue specific role (Eckmann et al. 2008). For example, aberrant NF-κB signalling in intestinal epithelial cells causes an inflammatory bowel like disease phenotype in mice, and is a critical regulator of epithelial integrity (Nenci et al. 2007), but NF-κB inhibition may have opposing roles in acute and chronic intestinal inflammation in different tissues (Eckmann et al. 2008). The role of NF-κB signalling in the AS gut epithelial cells would be an interesting comparison, since the gut epithelium is extremely important in the interaction between the mucosal immune system and gut microflora. NF-κB activation in the gut epithelium causes severe chronic intestinal inflammation in mice, apoptosis of colonic epithelial cells,

impaired expression of antimicrobial peptides and translocation of bacteria into the mucosa (Nenci et al. 2007). HLA-B27 may play a role in altering the microbiome in AS patients (Rosenbaum and Davey 2011), which may affect intestinal immune responses (Round and Mazmanian 2009). GWAS have shown a large overlap between AS and IBD associated loci, demonstrating a potential shared mechanism between NF-κB signalling and the regulation of barrier function in the gut. The potential pro- and anti-inflammatory role of NF-κB may prevent this pathway from being directly targeted as a therapeutic, since its interactions and feedback loops are extremely complex and yet to be fully understood in a tissue specific context.

Apoptosis related genes were greatly enriched in a number of analyses, and are also important components of the NF-κB signalling pathway. Apoptosis is an essential mechanism that prevents prolonged inflammation, and NF-κB has been demonstrated to have a pro-apoptotic role in neutrophils (Lawrence et al. 2001; Greten et al. 2007) and eosinophils (Kankaanranta et al. 2005) but has also been shown to inhibit apoptosis in macrophages (Park et al. 2005b) and therefore promote an inflammatory phenotype. The increase in B cell germinal centres with *TNFAIP3* deficiency may be due to the inability of A20 to mediate apoptosis due to elevated levels of NF-κB dependent anti-apoptotic proteins (Tavares et al. 2010). Apoptosis related pathways were only enriched in this analysis in monocytes and neutrophils, suggesting that the regulation of cell death by A20 may be cell type and context specific (Vucic et al. 2011).

NF-κB signalling must be able to readily integrate with other pathways to generate a coherent biological response to an environmental stimulus, and this cross talk is mediated by HAT and or HDAC signalling by modulating NF-κB or κB sites across the genome (Calao et al. 2008). Signalling molecules can mediate the accessibility of κB sites by acetylation, which may affect the downstream control of cell proliferation and apoptosis (Adcock 2006; Barnes 2006). NF-

κB signalling and HDAC type I signalling in $CD14^+$ monocytes and $CD16^+$ neutrophils suggests an important link between the pathways in these two cell types (Dai et al. 2005; Calao et al. 2008). $CD19^+$ B cells, although enriched for canonical NF-κB signalling, only demonstrated one differentially expressed HDAC (*HDAC10*), with no significance for *EP300* or *CREBBP* which are known to be NF-κB co-activators. This may suggest that the regulation of the NF-κB signalling by acetylation and deacetylation is not as prominent or important in $CD19^+$ B cells in AS pathogenesis. NK cells, $CD4^+$ and $CD8^+$ T cells also did not show differential expression of *EP300* or *CREBBP*, and also demonstrated no enrichment for canonical NF-κB signalling. HDACs may act to regulate *miR-146a*, as HDAC inhibitors were shown to increase *miR-146a* expression (Wang et al. 2013b). Other important pathways that cross-talk with NF-κB signalling include ubiquitin mediated proteolysis, which is thought to be regulated by TNFAIP3. TNFAIP3 cleaves the K63-linked ubiquitin chains for RIP1, TRAF6 and IKKγ and may inhibit signalling through TNFR complexes by binding to ubiquitin chains (Ma and Malynn 2012).

3.4.3 Potentially important long non-coding RNAs in AS

The role of lncRNAs in disease mechanisms cannot be overlooked, as lncRNAs have been recognised as important regulators of gene expression (Atianand and Fitzgerald 2014). This study replicated the findings of 3 known lncRNAs, *NEAT1*, *GAS5* and *PSORS1C3* in innate immune cell subsets, and also identified a large number of lncRNAs with unknown functions. *NEAT1* has previously been found to be up-regulated in MS patients (Santoro et al. 2016) and SLE patients versus healthy controls (Zhang et al. 2016) and was similarly up-regulated in the AS patients in $CD14^+$ monocytes and $CD16^+$ neutrophils. *NEAT1* is a key player controlling the formation of heterochromatin structures known as paraspeckles (Clemson et al. 2009), which may control target gene expression, specifically of

IL8 (Hirose et al. 2014). *GAS5* has been linked to cell cycle arrest (Kino et al. 2010; Mourtada-Maarabouni et al. 2010), and has also been found to regulate the inflammatory response of monocytes (Sun et al. 2017). The BXSB mouse strain (an important model of glomerulonephritis observed in SLE) demonstrated a down-regulation of *GAS5* (Haywood et al. 2006), which may be due to the potential loss of an Sp1 binding site from the promoter of *GAS5*. The down-regulation of *GAS5* in the AS patients may be due to decreased apoptosis, since this lncRNA may play a role in the cellular decision to enter the apoptotic pathway (Coccia et al. 1992). GAS5 has also been shown to act as a decoy glucocorticoid response element and modulating the transcriptional activity of the glucocorticoid receptor in mice (Kino et al. 2010). *PSORS1C3* was found up-regulated in AS patients in $CD16^{+}$ neutrophils, and is a known susceptibility gene for psoriasis (Chang et al. 2006). *SNHG5* was found differentially expressed in all cell types apart from NK T cells. Although not much is known of its function, *SNHG5* was found to demonstrate allele specific expression when mice strains were exposed to tunicamycin, a glycosylation inhibitor, causing the accumulation of misfolded proteins in the lumen of the ER and triggering a robust ER stress response (Chow et al. 2015). These results suggest that these lncRNAs may be important in regulating the immune response in AS and other immune mediated diseases, and could be used as markers of inflammation.

3.4.4 Towards the identification of biomarkers and new treatment options

Identifying novel associated changes in gene expression between AS patients and controls may pave the way to identifying effective biomarkers and new treatment strategies based on distinct molecular mechanisms. The expression of *TNFAIP3* is consistently down-regulated across AS patients and various autoimmune diseases, therefore presents itself as a possible therapeutic target. Current

therapeutic strategies based on antibodies or small molecules are more effective in reducing the function of a protein in contrast to restoring them, therefore the challenge is to devise an approach to increase levels of *TNFAIP3*. One approach is to inhibit the miRNAs that target *TNFAIP3* to increase its expression (Kim et al. 2012). Increasing *TNFAIP3* expression in mice was shown to improve colitis, with no effect on normal intestinal function, by inhibiting LPS-induced loss of the tight junction proteins and therefore improving intestinal barrier functions (Kolodziej et al. 2011). The important function of *TNFAIP3* in maintaining gut homeostasis again points to the importance of intestinal epithelial barrier integrity, and to investigate intestinal homeostasis in AS patients and how it is maintained through tight junction protein regulation. Targeting the microbiome may also be an efficient strategy if alterations in the intestinal flora are found in a subset of AS patients, as it appears that AS pathogenesis may be related to a balance between the host genetics, the intestinal microbiome and the regulation of the immune response.

The use of HDAC inhibitors in the treatment of cancers has suggested their efficacy in the treatment of inflammatory diseases. It has been demonstrated that HDAC inhibitors may exhibit anti-inflammatory effects in an adjuvant-induced arthritis model (Chung et al. 2003), and have also induced apoptosis in human eosinophils and neutrophils (Kankaanranta et al. 2010). HDAC inhibitors increased the expressions of *miR-146a*, and enhanced the negative regulation of IL-1 signalling (Wang et al. 2013b), demonstrating the potential anti-inflammatory effects of HDAC inhibitors and their use as an effective therapy. The expression of HDACs in this study were found up-regulated in AS patients in $CD14^+$ monocytes and $CD16^+$ neutrophils, but has some conflicting findings with a previous study measuring HAT/HDAC activity using specific assays (Toussirot et al. 2013). Although further investigation is required to

understand the precise role of HDAC signalling in AS aetiology, my data suggests HDACs may be a future therapeutic target in AS.

3.4.5 miRNAs in AS and immune mediated diseases

Understanding how miRNAs may be involved in AS pathogenesis is important due to their function as master regulators of gene expression, and may lead to novel therapeutic targets. Previous studies have produced differences in miRNA discovery, probably due to the small sample sizes and differences in AS patient inclusion criteria (Li et al. 2016). The most studied miRNA in AS is *miR-146a-5p*, due to its known role in regulating immunity and inflammation and the study of the variants that may be involved in its function (Niu et al. 2015; Xu et al. 2015; Chatzikyriakidou et al. 2010). It is encoded on chromosome 5q33, and mature miR-146a can bind to the 3' untranslated regions of many target mRNAs, including interleukin-1 receptor-associated kinase 1 (*IRAK-1*), *IRAK-2*, tumor necrosis factor receptor-associated factor 6 (*TRAF-6*) and other transcripts associated with inflammatory signalling. A polymorphism in *miR-146a*, rs2910164 was found to be associated with AS (Xu et al. 2015), and another study found an association with a SNP in *IRAK1*, a target of *miR-146a* (Chatzikyriakidou et al. 2010). Pro-inflammatory cytokines such as TNF-α and IL-1B, which were shown to be up-regulated in $CD14^+$ monocytes and $CD16^+$ neutrophils, have been reported to target *miR-146a* expression. Our results demonstrated a significant down-regulation of *miR-146a-5p* in $CD14^+$ monocytes of AS patients.

The *miRNA-146a* is involved in innate immunity by regulating the acute inflammatory response after pathogen recognition by TLRs on monocytes or macrophages (Taganov et al. 2006). Lv et al found that *miR-146a* was significantly down-regulated in the PBMCs of patients with intervertebral disc degeneration (IDD) and that over-expression of *miR-146a* was involved in decreasing the

levels of pro-inflammatory cytokines, meaning that the mRNA and protein levels of TRAF6 and NF-κB were down-regulated by *miR-146a* over-expression (Lv et al. 2017). The *miR-146a* was also found to be down-regulated in SLE patients in a number of studies (Tang et al. 2009; Lofgren et al. 2012; Ji et al. 2014). Diminished up-regulation of *miR-146a* was found in response to T cell stimulation in T regulatory cells (Tregs) of RA patients, with diminution of *miR-146a* expression observed particularly in patients with active disease, and correlated with joint inflammation. The Tregs of patients with active disease demonstrated a pro-inflammatory phenotype characterised by inflammatory cytokine expression, due to the activation of signal transducer and activator transcription 1 (STAT1), which is a direct target of *miR-146a* (Zhou et al. 2015). However, *STAT1* was not differentially expressed in our cohort. In contrast, *miR-146a* was found up-regulated in the synovial tissue and PBMCs of rheumatoid arthritis patients (Stanczyk et al. 2008; Pauley et al. 2008; Nakasa et al. 2008). The *miR-146a* is directly regulated by NF-κB levels, and when NF-κB accumulate, *miR-146a* levels increase, which upon processing and maturation will down-regulate *IRAK1* and *TRAF6* to reduce NF-κB levels in a negative feedback loop. This is important in the regulation of inflammation following activation of the innate immune response (Ma et al. 2011). The positive and negative feedback loops that operate in NF-κB signalling may be a reason for the variety of differences in *miR-146a* levels in our patient cohort. The enrichment of a large number of mRNAs and miRNAs in NF-κB signalling is substantial evidence to warrant further investigation in the role of these miRNAs and mRNAs in AS.

The *miR-146a* is also known to regulate *CXCR4* and *CXCR12*, chemokine receptors that are known to be aberrantly expressed in immune-mediated diseases (Werner et al. 2013). It also targets *ERBB4*, which is a receptor tyrosine kinase involed in promoting macrophage apoptosis and therefore limiting inflammation (Schumacher et al. 2017), as well as other genes known to be

involved in inflammation such as *CCL5* and *TLR4* (Marques et al. 2013; Tan and Farheen 2011). *CXCR4* and *CCL5* were found significantly down-regulated in AS patients $CD14^+$ monocytes, demonstrating a potential role of this miRNA in regulating a number of immune related mRNA targets.

Another miRNA involved in the NF-κB regulation was discovered as differentially expressed in $CD14^+$ monocytes. The *miR-16-1-3p* was up-regulated in AS patient monocytes without accounting for disease activity. Aberrant expression of *miR-16* was previously found in AS T cells rather than in monocytes, with higher expression in AS patients versus controls (Lai et al. 2013). Studies in related inflammatory diseases have found that *miR-16-1-3p* is also up-regulated in in mucosal tissues in UC (Wu et al. 2008) and Crohn's disease (Wu et al. 2008), plus in peripheral blood for UC (Paraskevi et al. 2012; Yang et al. 2013) and Crohn's disease (Iborra et al. 2013). The relevance of *miR-16* up-regulation in monocytes instead of T cells is not unsurprising, due to the importance of NF-κB in both innate and adaptive immunity. However, NF-κB signalling is complex, involving a large number of cross-related signalling pathways and a whole variety of miRNAs which may have not been detected in this study due to its limited power.

A novel finding from our study was *miR-100-5p*, which was found differentially expressed in AS $CD4^+$ and $CD8^+$ T cells in comparison to healthy controls. This miRNA is known to target *EGR2*, which is a transcription factor involved in T cell function in adaptive immune responses. Deletion of EGR2 and EGR3 in lymphocytes resulted in a lethal autoimmune syndrome with excessive serum pro-inflammatory cytokines but also impaired antigen receptor-induced proliferation (Li et al. 2012). *EGR2* was however not found differentially expressed in either $CD4^+$ and $CD8^+$ T cells. Mechanistic target of rapamycin (mTOR) is also a target of this miRNA, which is a component of two interacting complexes, mTORC1 and mTORC2, that regulate T cell lineage specification and

macrophage differentiation (Perl 2016). Genes relating to mTOR signalling were not found differentially expressed, leaving the significance of this miRNA yet to be uncovered. However, it could potentially act as a disease activity marker in AS T cells.

The *miR-409-3p* was found significantly up-regulated in high activity versus low activity AS patients and high activity AS patients versus healthy controls in $CD4^+$ T cells and was also identified in a previous study (Lai et al. 2013). It is known to target IFN-γ, a type II interferon, which is produced by lymphocytes activated by specific antigens or mitogens. *IFN-γ* was down-regulated in AS patients, suggesting this miRNA may have a role in blocking *IFN-γ* expression in severe AS patients (Wang et al. 2015). Some other interesting targets between high activity AS patients versus healthy controls found *miR-144-3p* and *miR-451a* to be differentially expressed. *MiR-144-3p* is known to target *TGFB1* which plays an important role in bone remodelling. It can also promote either T-helper 17 cells (Th17) or Treg lineage differentiation. The *IL6R* is a validated target for *miR-451a*, which is a known GWAS hit (Cortes et al. 2013).

3.4.6 Prioritised targets for future validation studies

This results chapter has provided evidence for the potential importance of a number of genes and signalling pathways in AS pathogenesis in a cell specific manner. Genes and miRNA differentially expressed within the NF-κB pathway require validation in a different $CD14^+$ monocyte cohort. This will aid in understanding how miRNAs, such as *miR-146a*, are regulating this important pathway. *TNFAIP3* is well known to be down-regulated in a number of autoimmune diseases as well as AS in this study and therefore should also be validated in further AS cohorts. A number of novel lncRNAs were identified across immune cell types, such as *NEAT1* in $CD14^+$ monocytes and $CD16^+$ neutrophils and *SNHG5* across cell types. Their functions in AS require

functional follow-up. The enrichment of HATs and HDACs in $CD14^+$ monocytes and $CD16^+$ neutrophils also required further validation in a separate AS cohort by RNA-seq or qPCR analysis.

3.4.7 Limitations and conclusions

This study is powered to detect cell specific changes in gene expression, therefore may reduce the noise most likely present in PBMC based expression studies. A limitations to this study is the small cohort size of between 10-15 AS patients, which may limit the power to detect some differentially expressed genes. The majority of AS patients included in the study exhibited active disease, but there are variations in the BASDAI score and other clinical features such as extra-articular manifestations and family history of autoimmune diseases. Patient heterogeneity may affect the outcome of many expression studies and GWAS, and may be controlled by extensive phenotyping. The 4 healthy controls used for the miRNA analysis were all female and HLA-B*2705 heterozygotes, which may have also have biased the results of this analysis. The samples in these analyses are not perfectly paired, therefore more targeted and specific experiments may be required to highlight the specific expression regulation between mRNA and miRNAs.

Cell type and location is also still an important factor to consider, and cells from within the synovium may more accurately represent the expression profile in AS, but these samples are often difficult or impossible to obtain. AS is believed to be a systemic disease, therefore PBMCs may still accurately represent these differences. Differentially expressed genes in peripheral blood cells in patients with PsA was found to be similar to their inflamed synovium, suggest that cells derived from blood samples can be a good substitute in the analysis of disease related gene expression changes (Dolcino et al. 2015). Another interesting relevant tissue that was demonstrated in recent epigenetic analysis

of AS associated loci is the intestinal epithelial cells (Li et al. 2017c), that have demonstrated extensive importance in maintaining barrier function in diseases relating to gut inflammation. These types of samples are difficult to obtain, but as the number of cell specific databases continue to increase, further cell type specific multi-omics data with continue to advance our understanding of AS pathogenesis.

We must also be aware that transcriptional changes may not be representative of changes at the protein level, due a variety of post-transcriptional modifications. Therefore, future proteomic based studies are required to validate the findings described in this chapter.

Chapter 4

Defining the chromatin accessibility and transcriptional landscape in AS

4.1 Introduction

4.1.1 The accessible genome

Around 80% of SNPs identifed through GWAS are located in non-coding portions of the genome (Hindorff et al. 2009), suggesting an important role of complex disease associated variants in gene regulation. This is substantiated by the observation that 60% of these non-coding variants are also located within DHS sites, and another 20% being in complete LD with DHS sites (Maurano et al. 2012). This enrichment of GWAS variants in open chromatin regions of the genome indicates that GWAS SNPs may be involved in altering the binding of TFs, as DHS sites reflect the occupancy of DNA binding proteins to create nucleosome free regions of DNA. The profiling of open chromatin regions and other genomic modifications, including TF binding, will enhance understanding of important regulatory regions of the genome. These will vary in different cell types and cell states.

4.1.2 Chromatin regulation and its role in disease

Analysis of how the regulatory chromatin landscape may change in response to inflammatory disease mechanisms may provide an approach to investigate the regulome for disease and biomarker discovery, in addition to the traditional transcriptomics methodologies. Variants associated with disease are often found in regulatory enhancers, with no definitive link to gene expression. Mapping chromatin accessibility and the disease specific regulome may provide additional information, also in the context of important TFs and pathways. An example of this is a study that identified AP-1 and Stat92E as key regulators of tumour development by the induction of JAK/STAT signalling by the use of ATAC-seq and FAIRE-seq (Davie et al. 2015). ATAC-QTLs can also inform our understanding of how variants may disrupt binding sites for transcription factors known to be important for cell differentiation (Cheng et al. 2016). Cheng et al describe a SNP, rs174556, disrupting the binding site for CCCTC-binding Factor (CTCF) and therefore modulating chromatin accessibility. Altering the binding of TFs and CTCF can cause changes in expression to nearby genes involved in increasing or decreasing disease risk.

There are a number of studies that have begun to investigate the feasibility of chromatin accessibility mapping using clinical samples. For example in chronic lymphocytic leukemia with the aim to understand cancer heterogeneity (Rendeiro et al. 2016). The applicability of chromatin profiling has also been demonstrated in SLE patients, and demonstrated increased genomic accessibility surrounding genes involved in naive $CD19^{+}$ B cell activation and the transcription factor binding sites that regulate B cell activation and differentiation (Scharer et al. 2016). Allelic bias was shown to be involved in determining gene expression by altering the chromatin state at promoters and

distal enhancers (Dixon et al. 2015), providing further evidence for the need of regulatory genome analysis in understanding disease mechanisms.

4.1.3 Sample processing and ATAC-seq

ATAC-seq is a relatively new and sensitive method to map chromatin accessibility with fewer cells than previous methods allowed, and large consortium projects such as Roadmap and BLUEPRINT have begun to use ATAC-seq for assaying rarer cell populations and in future extending to clinical samples (Bernstein et al. 2010; Adams et al. 2012). Single-cell ATAC-seq has also been used in primary leukaemia cells to capture distinct stages in cancer cell evolution (Corces et al. 2016). ATAC-seq in our cohort was carried out on primary cells isolated from either AS or healthy control blood samples on the same day as sample processing. This minimises changes in the epigenetic profile, in an attempt to capture the true variability between patients and controls. The ATAC-seq workflow from patient recruitment to data analysis is demonstrated in Figure 4.1, with steps A to C being completed on day one. ATAC-seq has a number of advantages over previous methods such as DHS, which is fuelling its use in clinical research projects. The number of cells required as input has been drastically reduced, from 10 million cells to roughly 50 thousand cells, enabling rarer cell populations to be investigated. Sample processing times are also reduced from roughly 4 days to half a day in comparison to DNase-seq (Buenrostro et al. 2013). These advantages along with the potential for single cell analysis, may create a role for future use of ATAC-seq in the clinic (Buenrostro et al. 2013), facilitating the identification of *cis*-regulatory elements to accompany traditional GWAS studies with the aim to relate non-coding, intergenic disease associations to regulatory elements (Scharer et al. 2016).

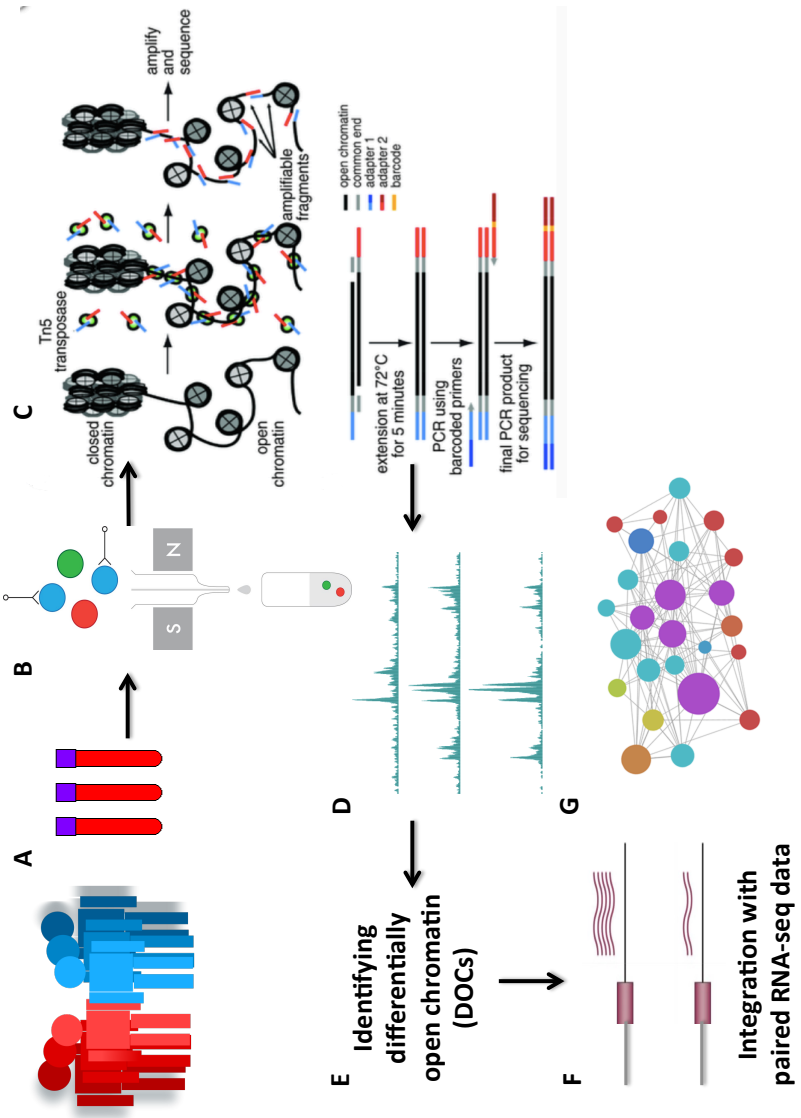


Figure 4.1: ATAC-seq workflow and analysis. (**A**) Recruitment and sample collection. (**B**) Magnetic-activated cell sorting (MACS, Miltenyi) to isolate specific cell populations. (**C**) ATAC-seq: transposition and library preparation. Adaptors are added with the transposition step and then completed with a 72°C amplification step. Sequencing indexes are then incorporated during the subsequent PCR. (**D**) Data processing, visualisation of read density distribution and the identification of open chromatin. (**E**) Identification of differentially open chromatin (DOCs) between patients and controls, and between cell types. (**F**) Integration with RNA-seq data from paired individuals to further identify which genes the DOCs may be regulating. (**F**) Pathway enrichment analysis and immune ontology analysis for the identification of important pathways and genes. Adapted from (Buenrostro et al. 2015).

4.1.4 TNF-α inhibitor therapy

TNF is a proinflammatory cytokine expressed predominately by activated macrophages and T cells, and is a potent activator of NF-κB - a transcription factor known for regulating many inflammatory genes. Enhanced production of TNF and TNF receptor (TNFR) 1 signalling is known to increase macrophage numbers and cytokine production in many common autoimmune or inflammatory diseases. TNF-α inhibitors (infliximab, eternacept, adalimumab and golimumab) have revolutionised treatment options for patients with AS and SpA, and work by binding to TNF-α to reduce its levels (Mitoma et al. 2016; Jin et al. 2010). They differ in their molecular structure, and also in their efficacy for different manifestations of SpA. Infliximab, adalimumab, golimumab and etanercept have excellent efficacy for articular involvement in SpA, but have very different effects on gut pathology. Adalimumab and infliximab are effective for IBD treatment, whilst etanercerpt is not (Song et al. 2008), but is effective at treating axial involvement.

Treatment response in AS was shown to be related to younger age, male sex, higher levels of inflammatory marker CRP. A higher BASDAI score related to treatment discontinuation (Arends et al. 2011). Treatment with TNF-α inhibitors generally leads to good improvement in clinical symptoms, CRP levels and MRI-detectable inflammation of the sacroiliac joints (Sieper and Poddubnyy 2016). Early treatment also reduced radiographic progression, suggesting that there may exist a window of opportunity to reduce inflammatory lesions leading to new bone formation (Maksymowych et al. 2013).

However not much is known about the pathways that regulate TNFR1 signalling, such as its cross talk with other relevant pathways that may become targets for next-generation biological therapies, such as the Toll-like receptor (TLR) signalling pathway. TLRs may bind to TNF and trigger downstream signalling

events such as apoptosis. In this study, samples from patients treated with 3 months of TNF-α inhibitors (with either adalimumab (ADA) or etanercept (ETA)) were collected. Samples were also collected from these patients before treatment commenced. ADA is a human monoclonal antibody, that blocks the soluble forms of TNF-α and cell surface TNF-α by inducing apoptosis. ETA is a dimeric fusion protein consisting of the extracellular ligand-binding portion of the tumor necrosis factor receptor (TNFR), which binds to the TNFR meaning that TNF-α is inactive. There are two previously published studies of relevance. The first examines the differences in gene expression between treatment responders and non-responders treated with ADA for 7 days in PBMCs (Dolcino et al. 2017). The second looks at gene expression changes in AS patients before and after treatment in PBMCs (Wang et al. 2017). Understanding the specific transcriptomic response to TNF inhibitors is therefore in its infancy, and the work presented in this chapter is the first such study to look at the effect of TNF-α inhibition in distinct purified immune cell populations. All four patients are treatment responders, but it is known that 40% of patients do not respond to TNF-α inhibition therapies (Roda et al. 2016). Therefore the identification of certain biomarkers that may be able to predict responsiveness and begin to explain possible molecular mechanisms are necessary. The effects of TNF-α inhibition goes well beyond the suppression of TNF alone (Horiuchi et al. 2010; Stone et al. 2004), and understanding these effects may aid in identifying biomarkers for treatment response and adverse effects.

4.2 Aims

This results chapter aims to determine differences in chromatin accessibility between AS patients and controls using ATAC-seq for different immune cell types. The goal is to define the AS-specific chromatin accessibility landscape that

can in turn identify likely regulatory DNA elements and inform understanding of disease pathogenesis and interpretation of GWAS variants. Specific aims are:

1. To identify differentially open chromatin regions between AS patients and controls for 3 different immune cell types.
2. To determine how chromatin accessibility changes in AS patients with active disease following anti-TNF treatment for 3 months across cell types.
3. To integrate analysis of differentially open chromatin with differential gene expression.

4.3 Results

4.3.1 Patient and control cohorts used for ATAC-seq analysis

Table 4.1 details the 5 AS patients recruited for the pilot study for ATAC-seq and total RNA-seq in three cell types through the GENExpress ID study. All 5 patients have high disease activity as described by the BASDAI criteria, with an average score of 7.2 (range 6.0-9.4), with a mean age of 32 years. One patient, GNOPBAS043, is significantly older at diagnosis than the other four, therefore could be classified as late onset AS (Hmamouchi et al. 2011; Montilla et al. 2012). Further details of the BASDAI classification for each patient can be found in Chapter 7. Healthy controls were also recruited between the ages of 30-50 (Table 4.2).

Patient ID	Sex	Age at Diagnosis	Disease duration (months)	BASDAI
GNOPBAS043	Male	52	36	9.4
GNOPBAS044	Male	14	300	6.4
GNOPBAS045	Male	27	120	6.5
GNOPBAS046	Male	31	36	7.7
GNOPBAS047	Male	35	9	6.0

Table 4.1: AS cohort 2 recruitment. The BASDAI or Bath Ankylosing Spondylitis Disease Activity Index consists of a 1 - 10 scale measuring discomfort, pain, and fatigue (1 being no problem and 10 being the worst problem) in response to six questions asked of the patient pertaining to the five major symptoms of AS. All patients have high disease activity and are from white British origin.

Control ID	Sex	Age
CTL1	Male	36
CTL2	Male	53
CTL3	Male	34
CTL4	Female	46
CTL5	Male	42
CTL6	Male	31

Table 4.2: Healthy control recruitment.

AS cohort characteristics

The patients were assessed by questionnaire on a number of questions relating to their lifestyle and health including smoking status, family history of AS, other autoimmune diseases and extra-articular manifestations (Table 4.3). Details of the patients current therapy, and the biologics therapy they were prescribed after their first visit is listed in Table 4.4. Initial recruitment criteria includes no previous biologics therapy. Four of the five patients have follow up samples

in each cell type, which are taken 3 months after starting on TNF-α inhibition therapy. Patient GNOPBAS047 is yet to commence biologic therapy. The global BASDAI score for each patient improved after 3 months on either etanercept or adalimumab, but the improvement was better for both patients on adalimumab (Table 4.4), where the disease activity decreased to low (lower than four). All 4 patients were treatment responders, and continued on the same therapy. Additional data for the BASDAI classification can be found in Chapter 7.

Table 4.3: AS cohort 2 metadata All patients are from a white British Background. PJI=Peripheral Joint Involvement

Patient ID	Smoker	Cigarettes/Day	Family history of AS	Other Autoimmune	Extra-articular
GNOPBAS043	Current	0.5	No	Sibling with RA	PJI
GNOPBAS044	No	-	Don't know	Don't know	None
GNOPBAS045	Previous	20	No	No	PJI
GNOPBAS046	No	-	Don't know	Don't know	Vertebral Fractures
GNOPBAS047	Current	10	Parent	Parent with RA	PJI

Table 4.4: AS cohort 2 therapy data. Inclusion critera for the first sample is biologic treatment naive. The follow up sample is a 3 months post anti-TNF therapy sample, of which 4 are available. The pre and post therapy global BASDAI score is also listed, which have decreased for all patients.

Patient ID	NSAIDs	DMARDs	Glucocorticoids	Biologics	Follow up sample	Pre/Post BASDAI
GNOPBAS043	Naproxen	-	-	Etanercept	Yes	9.4/7.8
GNOPBAS044	Naproxen	-	-	Adalimumab	Yes	6.4/1.5
GNOPBAS045	Arcoxia	-	-	Adalimumab	Yes	6.5/0.1
GNOPBAS046	Naproxen	-	-	Etanercept	Yes	7.7/4.7
GNOPBAS047	Naproxen	-	-	Etanercept	No	-

4.3.2 Differentially open chromatin regions between AS patients and controls

Quality Control of open chromatin regions

Forty-five primary human samples from 3 different immune cell types were processed using ATAC-seq, and the data was analysed using the pipeline created within the group, as previously described in Chapter 2. The number of paired-end reads varied between 40 to 200 million reads, and after filtering were reduced to between 35 to 120 million. The percentage of reads remaining after filtering varies from 70% to 30%, and is related to the percentage of mitochondrial reads before filtering. A general trend can be seen, with sets of samples with higher amounts of mitochondrial DNA producing a lower percentage of filtered reads (Figure 4.2(b) and (e)). The mitochondrial DNA ranged from 12% to 56% of the total reads. This is because it is easily accessed by the transposase since it is nucleosome free, and therefore has the opportunity to occupy a large percentage of the ATAC-seq reads. The number of peaks mainly ranged from 112 million to 200 million, with two $CD14^{+}$ AS samples exhibiting substantially higher numbers. These samples are the two GNOPBAS044 samples, which also seem to exhibit lower TSS enrichment (Figure 4.3(a)). The AS patients and controls demonstrate a slightly different trend in the number of filtered reads and the number of peaks, but all samples demonstrate a reasonable number of peaks expected for the number of filtered reads.

Open chromatin regions were detected using the peak calling software MACS2, which identified regions that are most accessible due to their larger amount of ATAC-seq reads owing to the increased integration of the Tn5 transposase. Between 112-270k peaks were identified at an FDR <0.01 (Figure 4.2(e)). Plotting the number of peaks against the number of filtered reads demonstrates a clear correlation at this peak calling threshold (Figure 4.2(f)). After filtering,

overlapping sample peaks were merged to create a set of union peaks that would be used to count differential openness for the remaining analysis (Figure 4.1). Samples were excluded if their Irreproducible Discovery Rate (IDR) between pseudo-replicates was low; CTL5 was excluded for $CD14^+$ and GNOPBAS046 for $CD4^+$ T cells. 2 samples were missing in the 3 month follow up samples due to an error in sequencing; GNOPBAS045 in $CD14^+$ monocytes, and GNOPBAS044 in $CD8^+$ T cells, therefore only 3 post treatment samples were available for these cell types. Signal to noise can also be measured by plotting the enrichment of ATAC-seq reads around the Transcription Start Site (TSS). The TSS plots for each cell type are demonstrated in Figure 4.3, and demonstrated good fold enrichment around the TSS, with CTL5 $CD14^+$ monocytes and GNOPBAS046 $CD4^+$ T cells showing least enrichment. GNOPBAS044 $CD14^+$ was kept for downstream analysis as its overlap with $CD14^+$ monocytes ENCODE DHS was shown to be above 80% (Figure 4.4(a)).

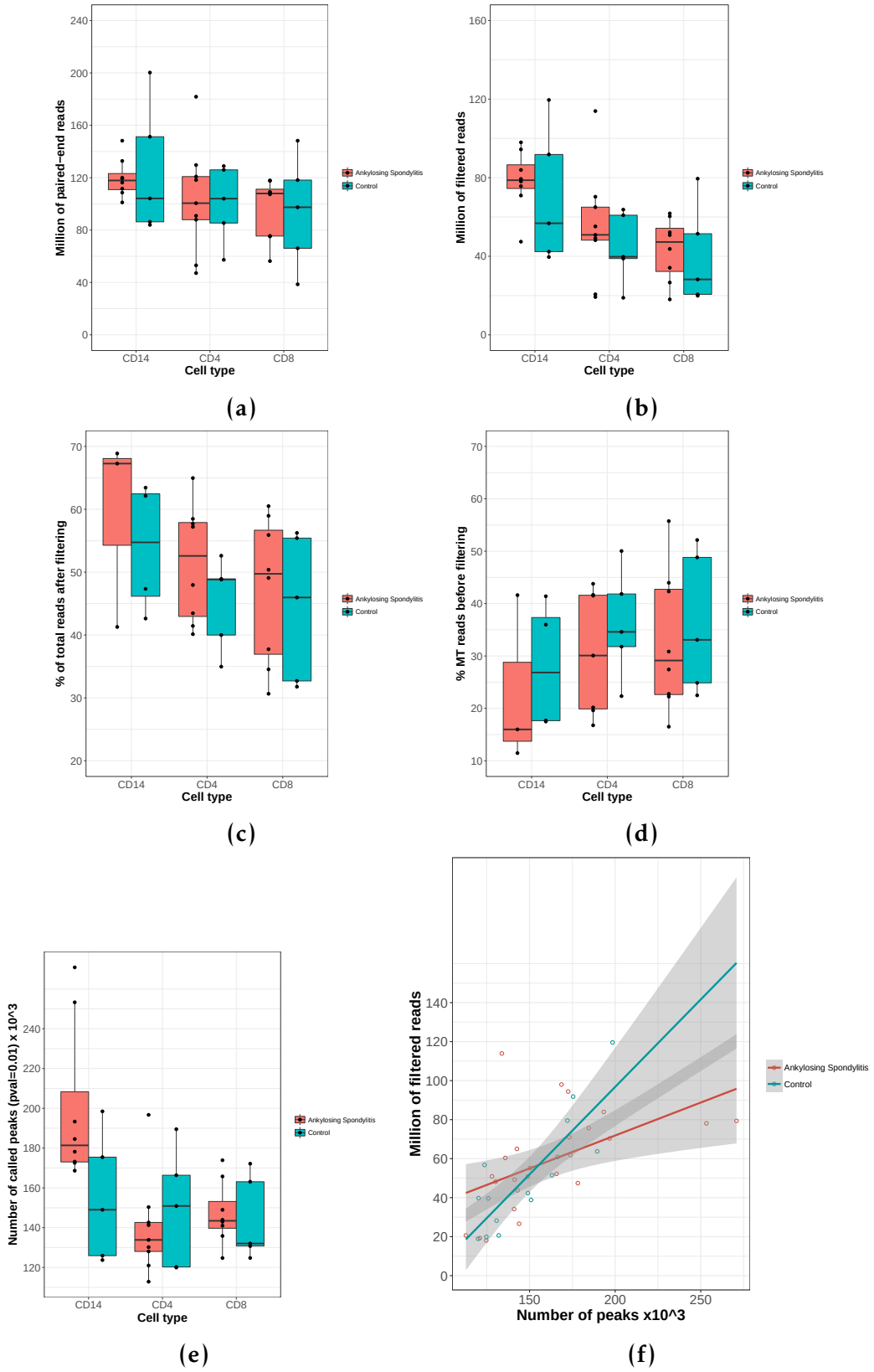
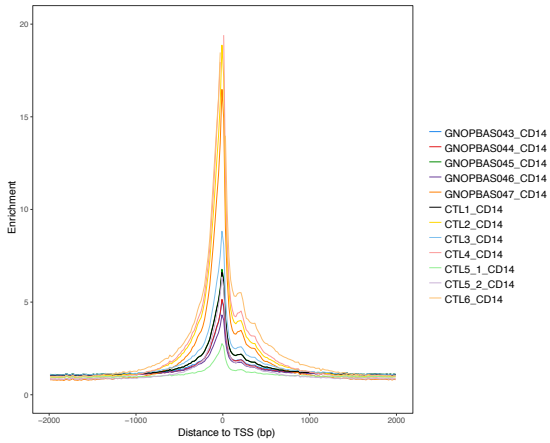
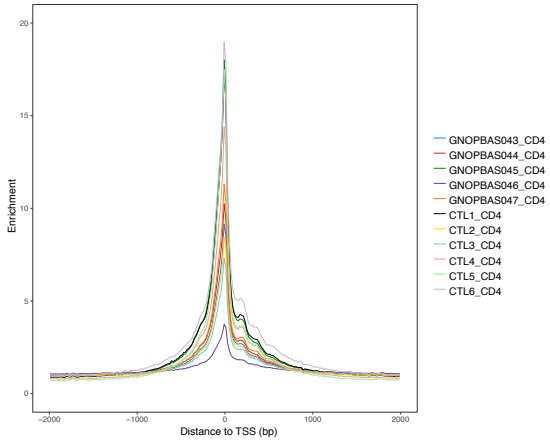


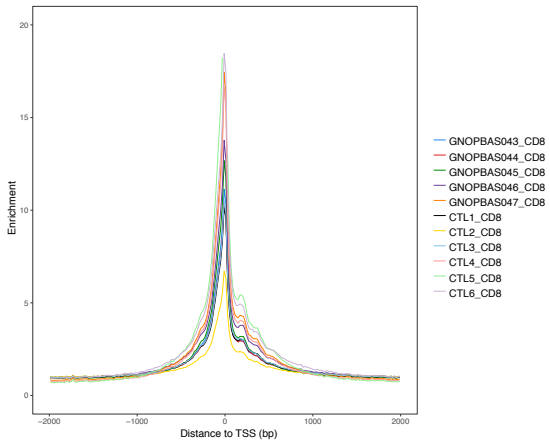
Figure 4.2: Quality Control of ATAC samples. Boxplots showing **(a)** the total number of paired-end reads. **(b)** the number of filtered reads. **(c)** the percentage of filtered reads from the total. **(d)** the percetage of mitochondrial reads before filtering. **(e)** the number of called peaks at P value <0.01. **(f)** Scatter-plot demonstrating the relationship between the number of filtered reads and the number of peaks.



(a)



(b)



(c)

Figure 4.3: TSS fold enrichment in controls and AS samples in (a) $CD14^+$ monocytes **(b)** $CD4^+$ and **(c)** $CD8^+$ T cells, with the majority of samples demonstrating strong enrichment. CTL5 $CD14^+$ (2 technical replicates) and AS046 $CD4^+$ demonstrated a fold enrichment of below 5, and therefore were excluded from the analysis. This is consistent with IDR analysis.

Peak master list generation

For each cell type, a merged master list was produced by using peaks that are found in at least 2 of the AS patients and healthy control samples. In $CD14^+$ monocytes a master list with 63200 peaks was produced, larger than the master lists for $CD4^+$ and $CD8^+$ T cells which produced similar lists of 46647 and 43554 peaks respectively. Creating a merged master list will create a stringent list of peaks, removing any background noise, as demonstrated in Figure 4.4(a) where the composite list of 3 samples created a much greater overlap, reaching almost 100% overlap with ENCODE DNase-seq data in the same cell type. Figure 4.4(a) also demonstrates the general quality of our data when overlapping with the gold standard DHS peaks from ENCODE, which can be improved by only using peaks found in more than one sample. The overlap with other ENCODE data, which are each demonstrated by a point in the graph is generally below 60%, demonstrating that the overlapping is greatly enriched within $CD14^+$ monocytes DHS sites.

In order to determine the extent of cell type specificity in the observed ATAC-seq peaks, unsupervised hierarchical clustering was performed on a master list including all samples for the 3 cell types. This revealed distinct clusters for $CD14^+$ monocytes and T cells. CTL5 for $CD14^+$ monocytes and AS046 for $CD4^+$ T cells again cluster separately here, reinforcing their poor quality. Batch effects (Figure 4.4(b)) will be corrected for in the differential analysis.

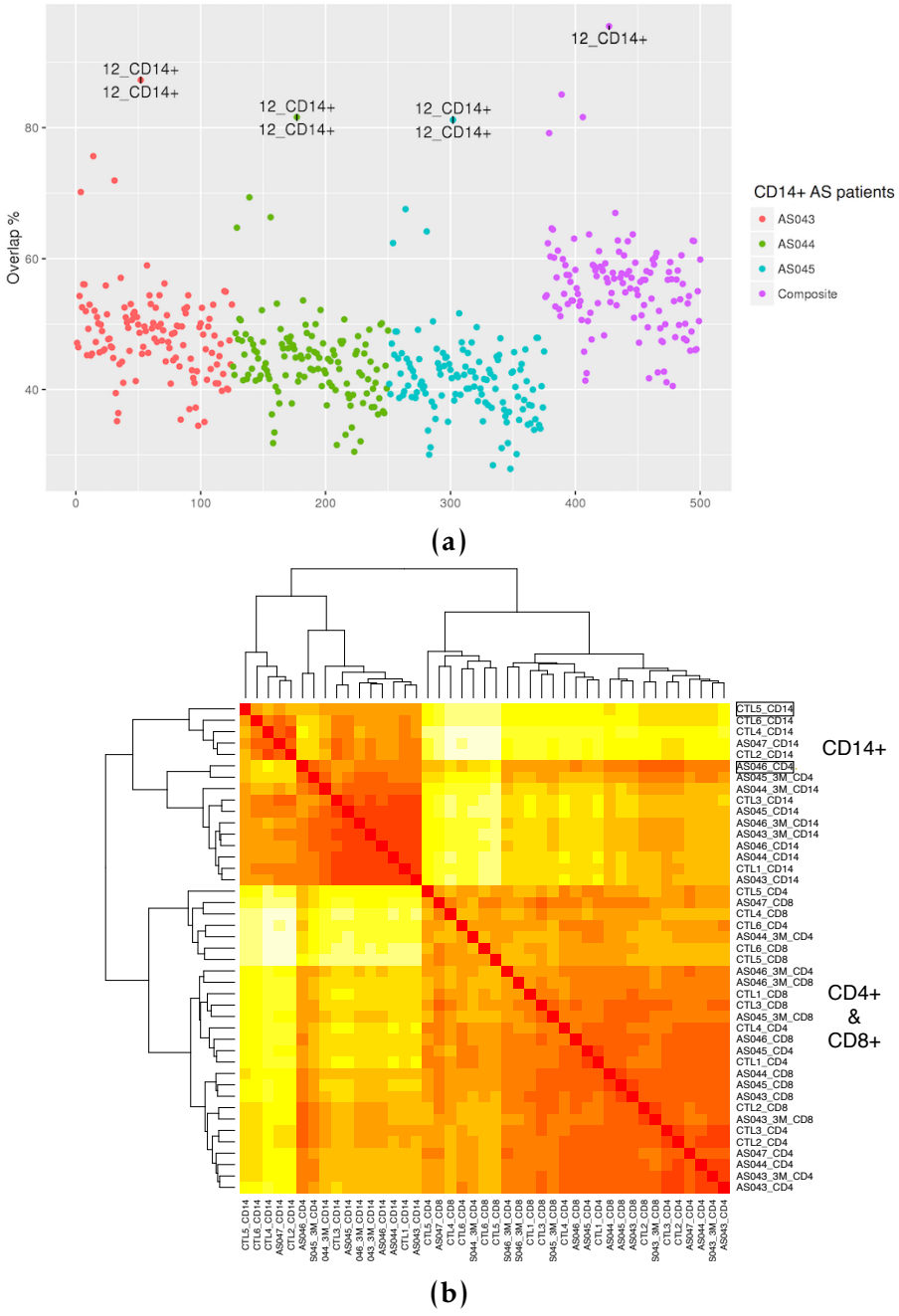


Figure 4.4: Merged peak master lists.(a) Overlap with ENCODE DNase-seq in 125 cell types for 3 AS CD14$^+$ samples, demonstrating that the percentage overlap increases when creating a merged master list, shown in purple. **(b)** Unsupervised hierarchical clustering heatmap created from a merged master list of all samples. CD14$^+$ monocytes are found distinct from T cells. The bad quality samples CTL5 CD14+ and GNOPBAS046 CD4$^+$ were also outliers in this plot. Sequencing batch effects can also be seen here.

Differential open chromatin analysis between AS patients and controls

Differential open chromatin analysis between AS patients and controls was performed using DESeq2, correcting for batch effects correlating with the date of sequencing, which could be seen on the unsupervised hierarchical clustering plot. Differentially open chromatin regions were identified for each cell type, the largest number being identified for $CD14^+$ monocytes, which also had the larger peak master list. The numbers of differentially open chromatin regions are listed in Table 4.5. An example of one of the top differentially open regions for $CD14^+$ monocytes is located in chromosome 5 (hg19 build, 131831656-131832235, FDR= 0.00016) and sits upstream of the gene *IRF1* as shown in Figure 4.5. There is also a second differentially open chromatin region around this location at the *IRF1* promoter (FDR= 0.007), both of which are differentially down-regulated in AS patients.

Cell Type	DOCs
	FDR 0.05/0.01
$CD14^+$	1815/141
$CD4^+$	29/0
$CD8^+$	178/42

Table 4.5: The number of differentially open chromatin regions between AS patients and healthy controls. An FDR cut off of <0.05 and 0.01 was used to define a differentially open peak.

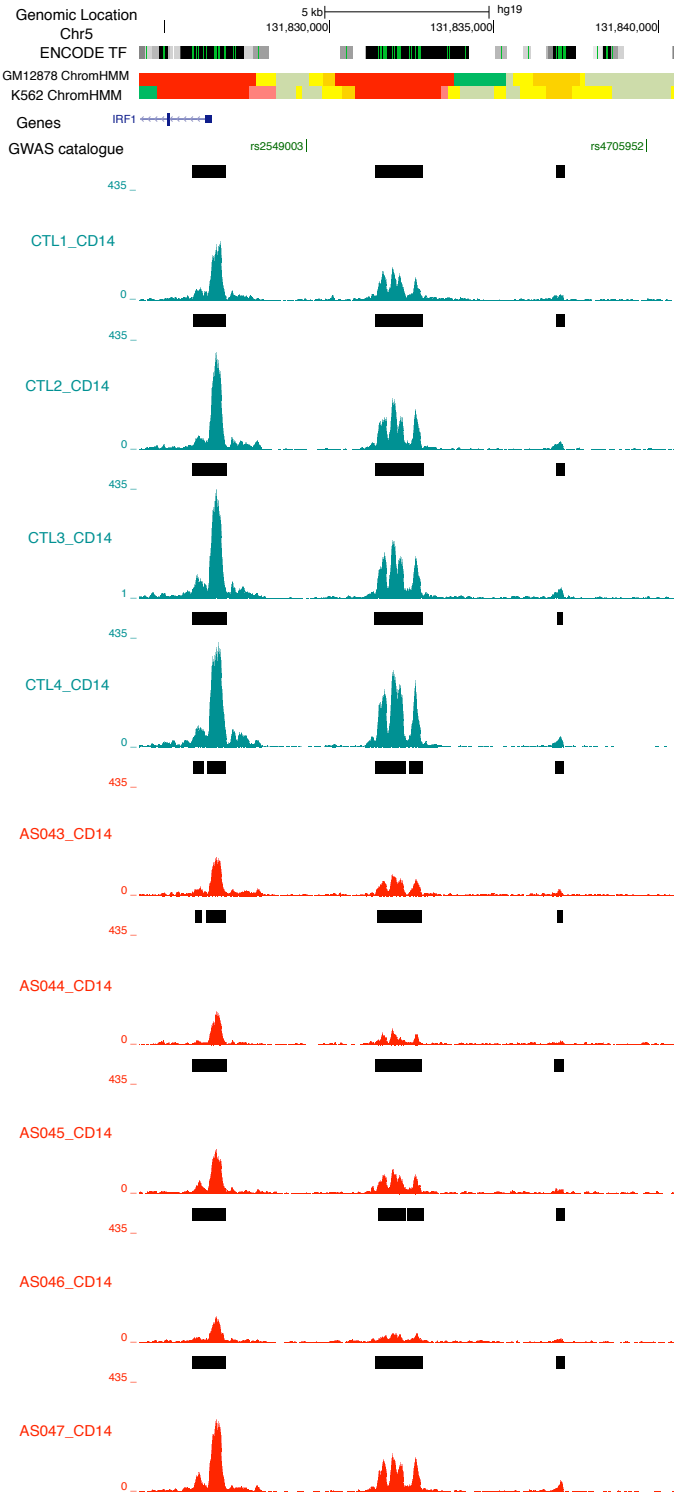


Figure 4.5: Differentially closed chromatin peaks in AS CD14$^+$ monocytes upstream of *IRF1* and at the *IRF1* gene promoter. UCSC Genome Browser view of the ATAC-seq read density distribution in CD14$^+$ monocyte healthy control samples (green) and AS patients (red). Both peaks are significantly down-regulated in AS patients (FDR= 0.00016 and 0.007 respectively, and overlap promoter predictions in GM12878 and K562 ChromHMM, along with overlapping with regions highly enriched for ENCODE TFBS.

RNA-seq analysis between AS patients versus controls

The RNA-seq samples from Table 4.1 were mapped using Tophat2 and gene counts generated using HTSeq-count as described in Chapter 2. All samples had reads mapping to known genes of over 10 million. GNOPBAS047 in $CD4^+$ T cells was excluded from the analysis due to low read counts (Appendix, Chapter 7, Figure 7.2).

Analysis carried out on this subset of patients revealed a large number of differentially expressed mRNAs and lncRNAs, which are described in Table 4.6. All samples were sequenced in the same batch, and age, sex or disease activity did not vary considerably between the samples, therefore the analysis was run without covariates using DESeq2 with approximately 25k transcripts. These RNA-seq results will be used specifically for integration with AS versus healthy controls ATAC-seq data. Refer to Chapter 3, Section 3.3.2 for a more comprehensive analysis of the transcriptomic landscape between AS patients and controls.

Cell Type	RNA	lncRNA
	FDR 0.05/0.01	FDR 0.05
$CD14^+$	2160/896	132
$CD4^+$	1411/587	98
$CD8^+$	1617/620	98

Table 4.6: The number of differentially expressed RNAs and lncRNAs between healthy controls and AS patients. An FDR cut off of <0.05 and 0.01 was used to define differentially expressed genes. Downstream analysis was carried out using the cut off of <0.05.

Integration of ATAC-seq and RNA-seq data

The significant differential open chromatin at an FDR <0.05 were each assigned to two genes by proximity, which were used as input for the comparison with the paired RNA-seq analysis results (FDR <0.05) from earlier in this chapter.

$CD14^+$ monocytes are significantly enriched in overlapping genes

$CD14^+$ monocytes exhibited an overlap of 437 genes with the differentially expressed genes. There were 1535 genes that were differentially expressed in RNA and not found differentially regulated in the ATAC-seq list of genes, and 2331 genes that were unique to ATAC-seq, against all genes present in both RNA-seq and ATAC-seq. These 437 genes were significantly more likely to be enriched than by chance (P= 2.2x10^{-16}, Fishers test).

RGS1 is differentially expressed in AS T cells

$CD8^+$ T cells had an overlap of 22 genes between the differentially expressed ATAC-seq and RNA-seq genes, with 212 genes that were present in just the ATAC-seq, and 1446 in the RNA-seq against all genes analysed in both datasets (15634). However, the enrichment of these genes was not found to be significant (P= 0.90, Fishers test). $CD4^+$ T cells demonstrated just one overlapping gene, regulator of G-protein signalling 1 (*RGS1*), which has been associated with multiple autoimmune disorders including type I diabetes. This gene was also found in $CD8^+$ T cells but not in $CD14^+$ monocytes, suggesting it may be a T cell specific gene. Promoter Capture Hi-C data (Mifsud et al. 2015) demonstrates a weak interaction with the ATAC-seq peak chr1:192578068-192579250 and the gene promoter of *RGS1*, which however does not reach significance (CHiCAGO of 3.22). It appears that the power to detect significant open regions in the T lymphocytes subsets is not sufficient, and the addition of more samples may improve the identification of differentially open chromatin regions. These results

were then used to intersect with known AS reported GWAS genes; 2 GWAS genes in $CD14^+$ were identified at *ZMIZ1* and *ETS2*. However, these peaks do not overlap with any AS associated SNPs identified by GWAS.

Pathway enrichment analysis in overlapping genes in $CD14^+$ monocytes

$CD14^+$ monocytes in controls versus AS patients had a significant number of overlapping genes to carry out pathway enrichment analysis. Differentially expressed genes and genes annotated close to the differentially open chromatin regions were analysed with the Hallmark gene set, using all annotated genes in the ATAC-RNA differential analysis as a background. Table 4.7 lists the pathways and genes that are found enriched at FDR <0.05. Differentially expressed genes are highly enriched in inflammatory response pathways, highlighting the importance of NF-κB signalling and the UPR.

Pathway	Genes
Genes regulated by NF-kB in response to TNF	*BCL3, SAT1, B4GALT1, SQSTM1, ATF3 CD44, PNRC1, NFIL3, KLF10, ZFP36, IRS2 ETS2, SLC2A3, MXD1, MAP3K8, NFKBIA SPHK1, TRIB1, NINJ1, JUNB, REL, FUT4 PHLDA2*
Genes up-regulated by activation of the PI3K/AKT/mTOR pathway	*FASLG, SQSTM1, CDKN1B RAC1, GRB2, MYD88, RIPK1 ACTR2, UBE2D3, ARF1, SMAD2, MAPK1*
Genes up-regulated in response to TGFB1	*KLF10, SPTBN1, UBE2D3, JUNB PPM1A, ID1, TGFBR1*
Genes up-regulated during unfolded protein response	*ATF3, NOLC1, RRP9, EDEM1, EIF4A1 YWHAZ, HYOU1, EIF2AK3, SLC30A5*

Table 4.7: Enriched pathways for the ATAC-seq differential open chromatin regions and RNA-seq differentially expressed genes in $CD14^+$ monocytes. FDR <0.05.

4.3.3 Response to anti-TNF therapy in AS patients analysed by ATAC-seq

Quality control

The quality control for the TNF-α inhibition samples are described collectively in Figure 4.2, and are also shown in the unsupervised hierarchical clustering plot in Figure 4.4(b). GNOPBAS044 3M was excluded for $CD14^+$ monocytes due to a low IDR, along with GNOPBAS045 3M and GNOPBAS046 3M for $CD4^+$.

Differential open chromatin analysis pre and post 3 month anti-TNF treatment

Master lists were created for AS patients and their paired 3 month TNF-α inhibition therapy samples, enabling a paired cross time point comparison. 43089, 38663 and 39402 peaks were found for $CD14^+$ monocytes, $CD4^+$ T cells and $CD8^+$ T cells respectively. Differential open chromatin analysis using DESeq2 was carried out between AS patients and their paired samples after 3 months treatment with anti-TNF therapy, taking account of the paired samples using the multifactor design. No differentially open chromatin regions were found in $CD14^+$ monocytes and $CD4^+$ T cells, which may relate to the quality of the samples, but $CD8^+$ T cells demonstrated a large number of differentially open regions; 116 passing an FDR <0.05. The top differentially open chromatin region was found in chromosome 5 (chr5:40410331-40410886), which is found close to the gene *PTGER4*, reported as an AS GWAS associated gene (Cortes et al. 2013).

RNA-seq paired expression analysis with 3 months anti-TNF therapy

Differential expression analysis was performed using the multifactor design in DESeq2, taking into consideration paired samples between treatment naive AS patients and their paired post 3 month TNF-α inhibition therapy samples (Table

4.8). All of the patients exhibited reduced global BASDAI scores after three months of treatment (Table 4.4), and were described as disease responders. Disease activity was not corrected for in the model.

Volcano plots of the up and down-regulated genes for each cell types are shown in Figure 4.6. The most interesting genes in $CD14^+$ monocytes include the up-regulation of GTPase IMAP Family Member 5 (*GIMAP5*), a gene involved in apoptosis and NF-κB signalling, and the down-regulation of Kruppel Like Factor 11 (*KLF11*), also involved in apoptosis and TGF-β signalling. The most significant gene in $CD4^+$ T cells following treatment was the down-regulation of *CXCR4*, a chemokine receptor involved in the inflammatory response. $CD8^+$ T cell exhibited an up-regulation of *CXCR6*, another chemokine receptor and *EOMES*, a TF involved in the differentiation of $CD8^+$ T cells during the immune response.

Cell Type	mRNA	lncRNA
	FDR 0.05/0.01	FDR 0.05
$CD14^+$	571/190	13
$CD4^+$	45/21	2
$CD8^+$	288/114	12

Table 4.8: The number of differentially expressed mRNAs and lncRNAs between AS treatment naive and post 3 months anti-TNF therapy. An FDR of <0.05 and 0.01 was used to define differentially expressed genes. Downstream analysis was carried out using FDR <0.05.

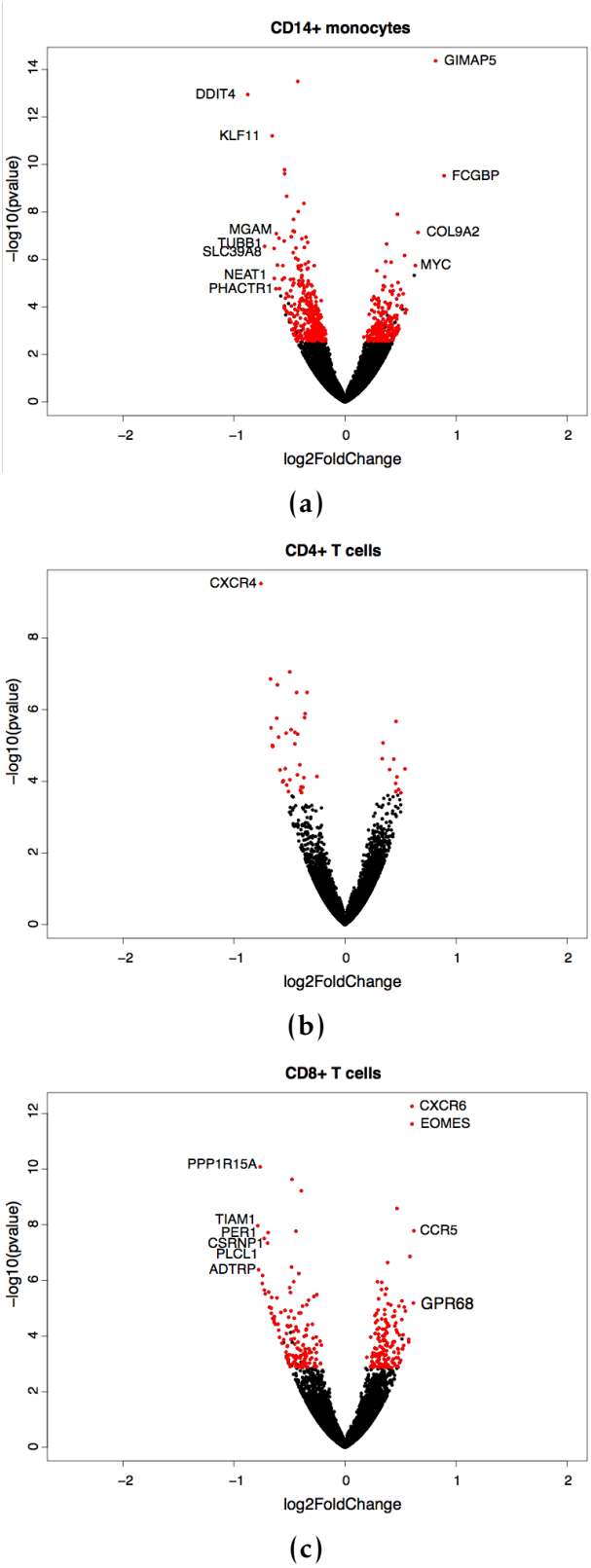


Figure 4.6: Volcano plots of differentially expressed genes in anti-TNF treated AS patients versus naive AS patients. The red points demonstrate genes differentially expressed with an FDR <0.05 in **(a)** $CD14^+$ monocytes **(b)** $CD4^+$ T cells **(c)** $CD8^+$ T cells.

Differentially expressed lncRNA following anti-TNF treatment

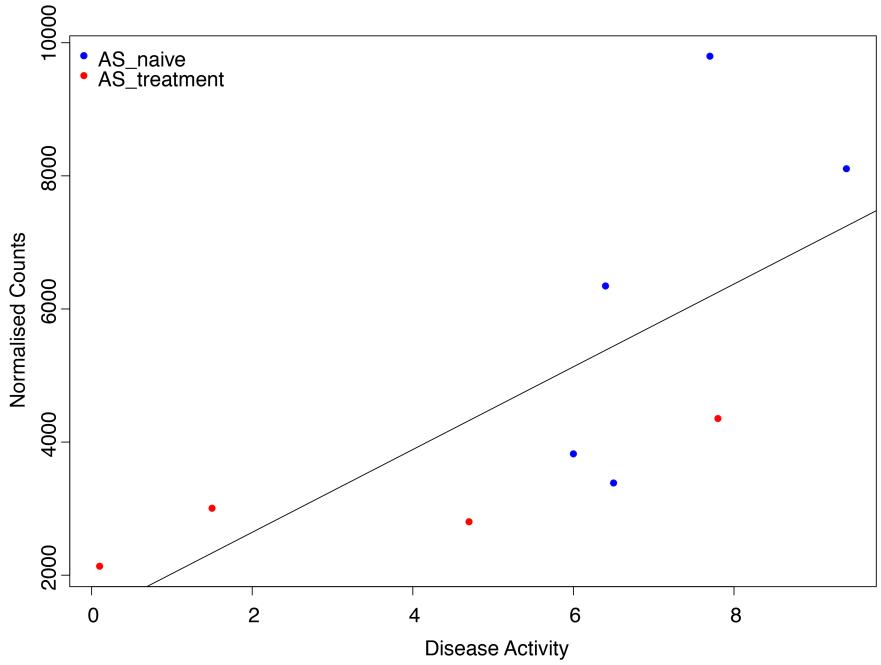
All differentially expressed lncRNAs at an FDR <0.05 are shown in Table 4.9. Many lncRNAs have limited functional information, therefore it is challenging to understand how these lncRNAs may be regulating genes related to AS. The two differentially expressed lncRNAs for $CD4^+$ do not have a known function. The differentially expressed genes and lncRNA may be described as markers of therapy response.

$CD14^+$	$CD4^+$	$CD8^+$
AC013394.2	CTB-31O20.2	MKLN1-AS1
MIR22HG	RP11-489E7.4	HCP5
NEAT1		NEFL
SCARNA2		RP11-94L15.2
RP11-473I1.10		RP11-489E7.4
HCP5		RP11-861A13.4
AP003774.6		LINC00402
HOTAIRM1		SNHG15
TUG1		AC084018.1
RP11-473I1.9		CTD-3184A7.4
SNHG15		RNU12
LINC00963		AC013394.2
RP11-1002K11.1		

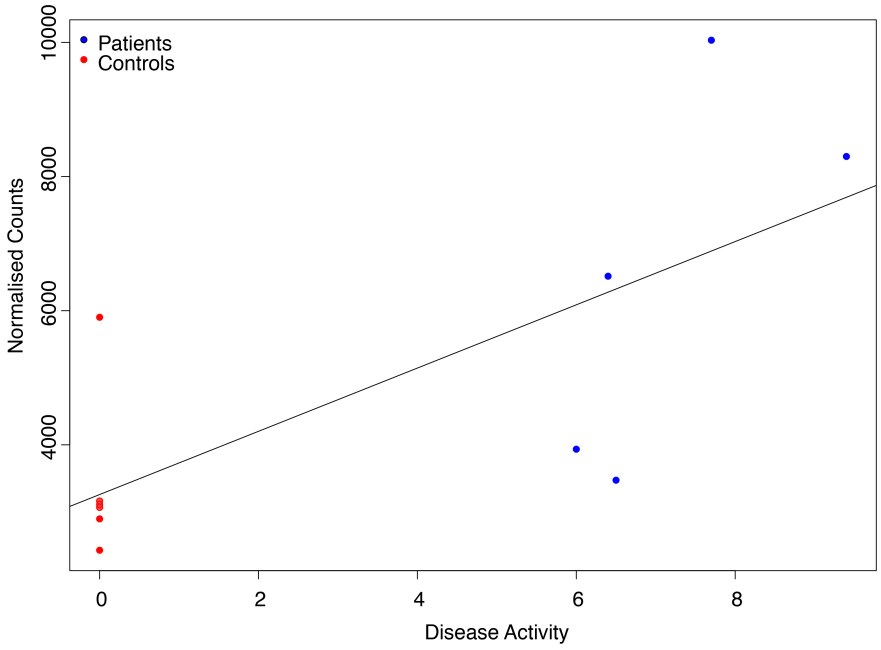
Table 4.9: Differentially expressed lncRNAs in anti-TNF treated AS patients versus AS treatment naive patients. FDR <0.05.

There are three overlapping differentially expressed lncRNAs between $CD14^+$ and $CD8^+$ T cells: *AC013394.2* and *SNHG15* which are both down-regulated in the two cell types in response to treatment and *HCP5* which is up-regulated in both.

Interrogation of differentially expressed lncRNAs in $CD14^+$ monocytes revealed a number of interesting targets including the up-regulation of *SCARNA2* in the treatment group. It is associated with the Cajal body, an organised nuclear structure which has a role in epigenetic control (Morris 2008). HOX antisense intergenic RNA myeloid 1 (*HOTAIRM1*) and *NEAT1* were both down-regulated in treated $CD14^+$ monocytes. *HOTAIRM1* has an emerging role in the innate immune response (Elling et al. 2016), and its down-regulation has been shown to attenuate myeloid differentiation (Zhang et al. 2014a). *NEAT1* has previously been found to be up-regulated in MS patients (Santoro et al. 2016) and SLE patients versus healthy controls in relation to disease activity (Zhang et al. 2016). Plotting the normalised counts against disease activity in our pre-treatment and treated AS patients revealed a strong correlation of higher counts with higher disease activity (Figure 4.7(a)). *NEAT1* expression in the $CD14^+$ healthy controls versus AS patients was marginally significant (FDR=0.056), but plotting the normalised counts confirms higher expression of *NEAT1* with high disease activity (Figure 4.7(b)).



(a)



(b)

Figure 4.7: Scatterplot of normalised counts showing *NEAT1* expression in CD14$^+$ monocytes in relation to disease activity in **(a)** Treated versus untreated AS patients (r^2=0.50, *P*=0.031). **(b)** healthy controls versus AS patients (r^2=0.44, *P*=0.048).

RNA-seq pathway enrichment in CD14$^+$ monocytes, CD4$^+$ and CD8$^+$ T cells following anti-TNF treatment

Pathway enrichment analysis of the differentially expressed genes between treatment naive AS patients and their paired follow up samples at FDR <0.05 in CD14$^+$ monocytes, CD4$^+$ and CD8$^+$ T cells, was analysed using XGR (Fang et al. 2016). Gene set enrichment analysis was analysed using the Hallmark gene sets by MSigDB, and pathway enrichment analysis using Canonical, KEGG, REACTOME and BioCarta gene sets. Pathway enrichment analysis was not carried out for the control samples versus AS patients as a more accurate pathway enrichment analysis is shown in Chapter 3.

Comparing AS patients with their post 3 month TNF-α inhibition therapy sample identified the top enriched pathway across cell types as genes regulated by NF-κB in response to TNF. All hallmark gene sets are listed in Table 4.10, and Canonical, KEGG, REACTOME and BioCarta results are shown in the Appendix (Chapter 7, Tables 7.4, 7.5 and 7.6). Plotting the differential gene expression data onto a manually curated KEGG TNF signalling pathway demonstrates some differences between the 3 cell types in the genes that are significantly differentially expressed in response to TNF-α inhibition. CD4$^+$ T cells only demonstrate one significant differentially expressed gene in this pathway (*PIK3R1*, FDR <0.05). Figures 4.8 and 4.9 illustrate the output for CD14$^+$ monocytes and CD8$^+$ T cells. The genes are coloured by the log2 fold change, with those that pass the FDR <0.05 highlighted in bold. CD14$^+$ monocytes contain 5 genes that pass significance, and CD8$^+$ T cells 9. Only one gene overlaps between the two cell types; *CASP8* expression is increased significantly in both with treatment. There are a number of pathways that cross-talk with the TNF signalling pathway that contain these genes, such as Apoptosis (*CASP8*), MAPK signalling (*MAPK1*, *MAPK14*) and P13K-Akt signalling (*PIK3CB*, *PIK3R1*). CASP8 is also known to have non-

apoptotic functions such as regulating cytokine production and interactions with inflammasomes.

$CD14^+$ monocytes also had significantly down-regulated *ICAM1*, which is a cell adhesion molecule found on the surface of immune cells. Cell adhesion molecules (CAMs) have roles in cell proliferation, differentiation, motility, apoptosis and tissue architecture. A number of intracellular signalling molecules and transcription factors were also down-regulated in $CD8^+$ T cells, including *JUN*, *SOCS3*, *JUNB*, *FOS*, but there was an up-regulation of *CCL5*, a chemokine receptor involved in leukocyte recruitment. *TNFRSF1A*, was also found to be up-regulated in both cell types, but only significantly in $CD8^+$ T cells, and is a GWAS hit for AS (Cortes et al. 2013). This is in contrast with the study by Wang et al, which found decreased levels of *TNFRSF1A* after treatment in PBMCs (Wang et al. 2017). *MAPK14*, which was found up-regulated in $CD8^+$ T cells was also recently described as being associated with SpA in a family based study (Costantino et al. 2017).

Some AS GWAS genes were found differentially expressed after treatment in a cell type specific manner that may have been undetected by previous PBMC studies. Zinc Finger MIZ-Type Containing 1 (*ZMIZ1*) was found significantly down-regulated after treatment (FDR= $1.72\text{x}10^{-05}$) in $CD14^+$ monocytes only, its down-regulation may be involved in the transcriptional activation of a subset of NOTCH1 target genes including *MYC*. *ETS2* and *FCGR2A* were also significantly down-regulated in $CD14^+$ monocytes, and *PTGER4*, *IL6R* and *CARD9* significantly up-regulated. In $CD4^+$ and $CD8^+$ T cells, *PTGER4* was significantly down-regulated in contrast with $CD14^+$ monocytes, demonstrating the importance of isolating distinct immune cell populations. *IL6R* was found significantly down-regulated post-treatment in $CD8^+$ T cells, consistent with a previous study (Wang et al. 2017).

There was a general dampening of immune-related genes when treated with TNF-α inhibitors, as demonstrated in the NF-κB signalling KEGG pathway in Figures 4.10, 4.11 and 4.12. Figure 4.10 demonstrates gene expression profile in AS CD14$^+$ monocytes in relation to healthy controls from Chapter 3. Figure 4.11 demonstrates the gene expression profile in AS anti-TNF treated CD14$^+$ monocytes against healthy controls and now shows no differentially expressed genes, with *TNFAIP3*, *NFKBIA*, *IL1B* and *IRAK1* increasing in expression in the treated AS samples, and *PTGS2* and *IKBKB* no longer reaching significance. Figure 4.12 demonstrates the gene expression profile in treated AS patients against naive AS patients, where again nothing is significant, apart from *PLCG1*, which is up-regulated in the treated AS samples. The protein encoded by this gene plays an important role in the regulation of intracellular signalling cascades, and has been described as a candidate gene in IBD (Yuan et al. 2017). This analysis demonstrates the effects of blocking TNF-α, but also shows that the majority of genes found in the canonical NF-κB signalling pathway between treated and untreated AS patients in CD14$^+$ monocytes do not differ significantly.

Table 4.10: Hallmark gene sets enriched in 3 cell types comparing AS patients with their post 3 month anti-TNF therapy samples. All pathways shown have an FDR <0.05.

CD14$^+$	**CD4$^+$**
Genes regulated by NF-κB in response to TNF	Genes regulated by NF-κB in response to TNF
Genes up-regulated during UPR in the ER	Genes defining inflammatory response
A subgroup of genes regulated by MYC	Genes up-regulated through activation of mTORC1 complex
Genes up-regulated in hypoxia	Genes mediating apoptosis
Genes up-regulated in response to TGFB1	Genes up-regulated in hypoxia
Genes up-regulated by STAT5 by IL2 stimulation	Genes up-regulated in response to IFNG
	Genes up-regulated by STAT5 by IL2 stimulation
CD8$^+$	
Genes regulated by NF-κB in response to TNF	Genes mediating apoptosis
Genes up-regulated in response to IFNG	Genes up-regulated by PI3K/AKT/mTOR pathway
Genes up-regulated by STAT5 in response to IL2 stimulation	
Genes up-regulated by IL6 via STAT3	

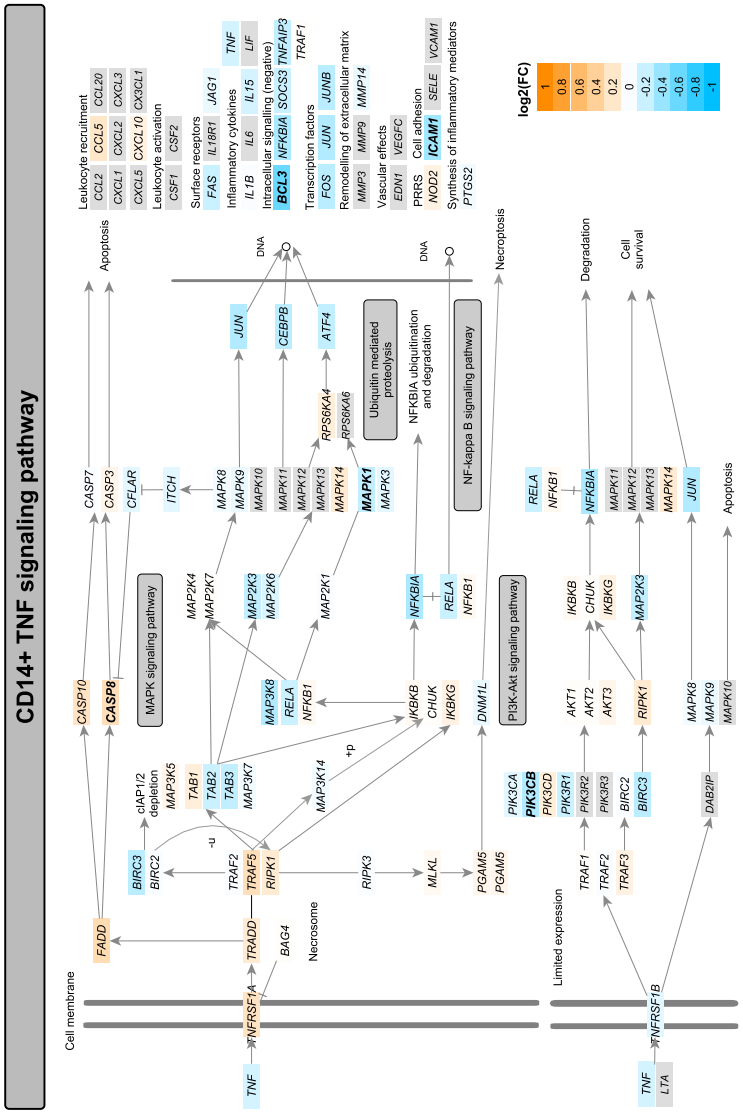


Figure 4.8: CD14$^+$ monocytes show attenuation in TNF-α signalling pathways following 3 months TNF-α inhibitor therapy, compared with paired untreated AS patients. Genes are coloured by their log2 Fold Change, with significant genes at FDR <0.05 highlighted in bold.

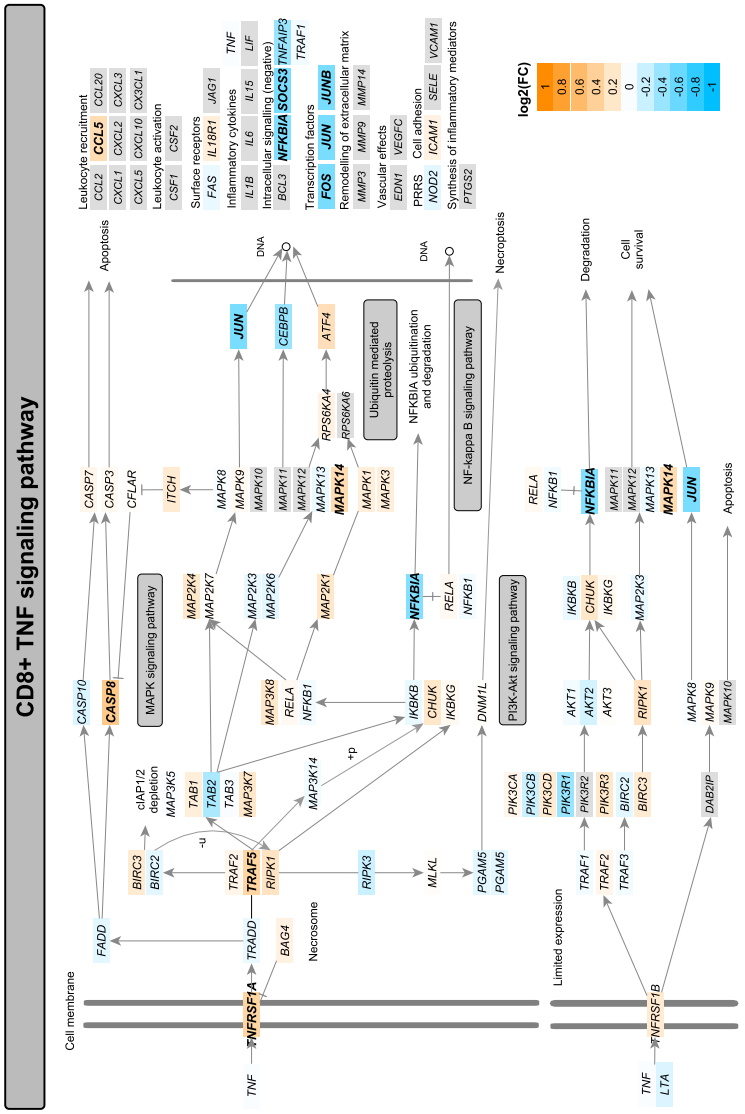


Figure 4.9: CD8$^+$ T cells show attenuation in TNF-α signalling pathways following 3 months TNF-α inhibitor therapy, compared with paired untreated AS patients. Genes are coloured by their log2 Fold Change, with significant genes at FDR <0.05 highlighted in bold.

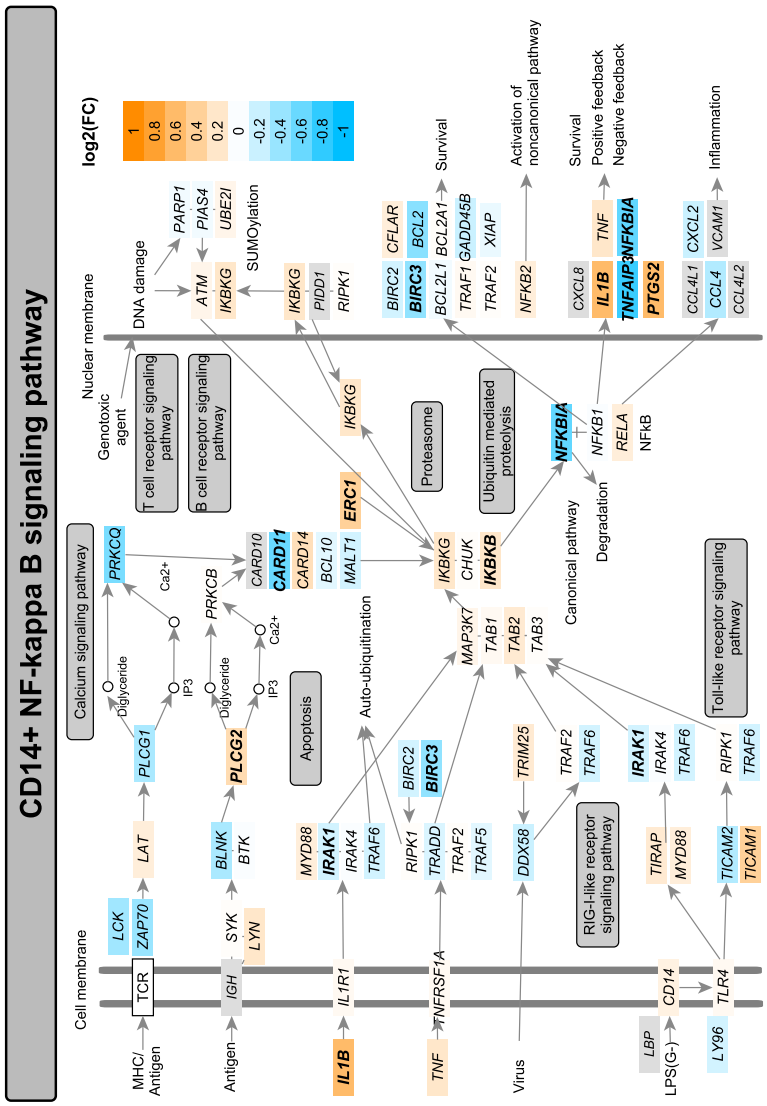


Figure 4.10: CD14$^+$ monocytes comparing AS patients and healthy controls in the NF-κB signalling pathway. Genes are coloured by their log2 Fold Change, and significant genes are in bold.

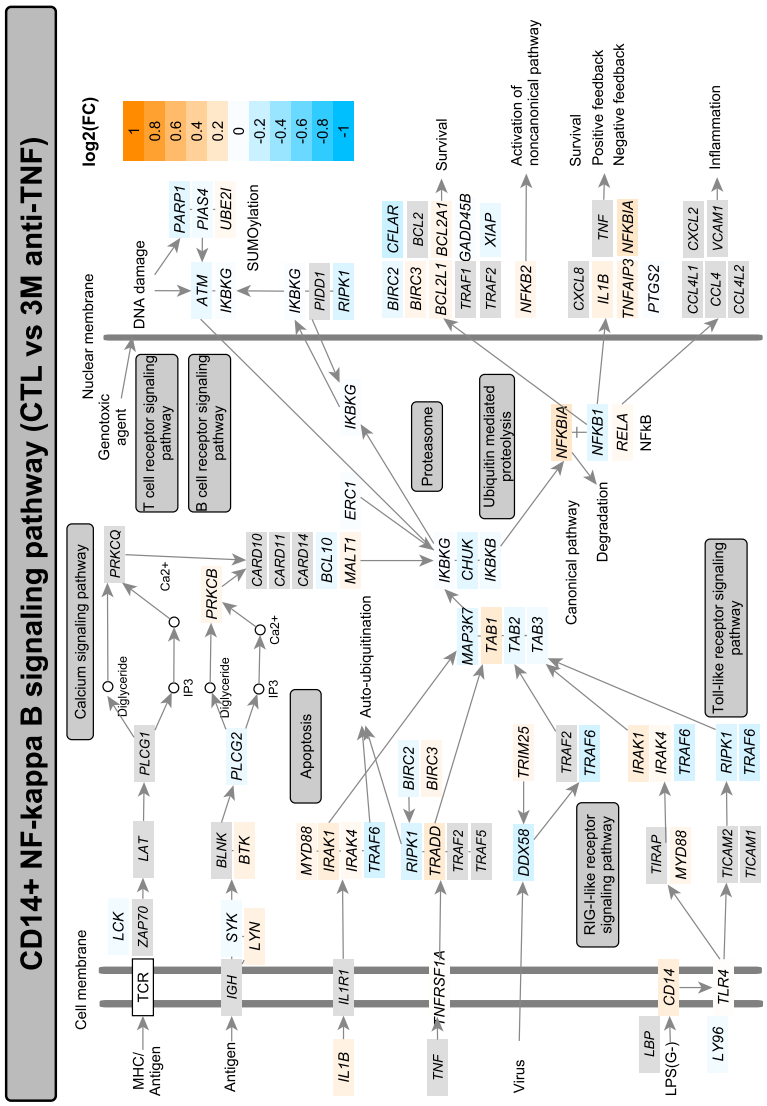


Figure 4.11: CD14$^+$ monocytes comparing AS treated patients and healthy controls in the NF-κB signalling pathway. Genes are coloured by their log2 Fold Change, and significant genes are in bold.

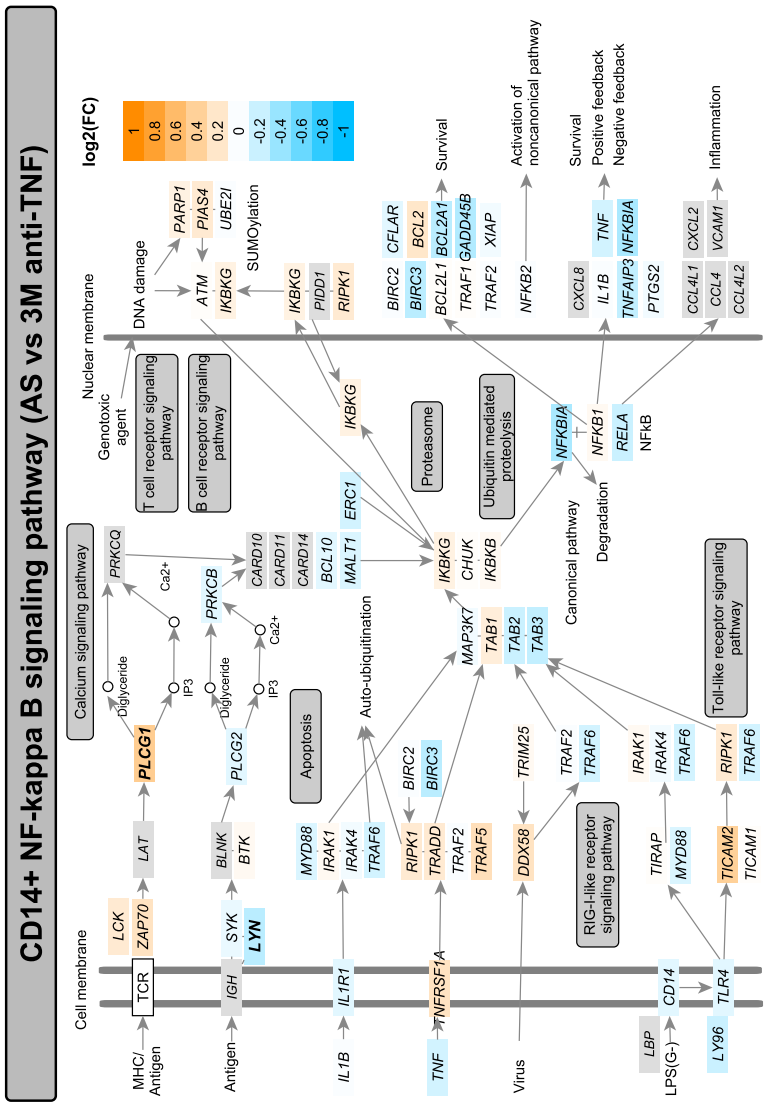


Figure 4.12: CD14$^+$ monocytes comparing AS treated and untreated patients and healthy controls in the NF-κB signalling pathway. Genes are coloured by their log2 Fold Change, and significant genes are in bold.

Integration of ATAC-seq and RNA-seq data reveals regulatory changes in response to treatment

Significant differentially open chromatin regions were identified in $CD8^+$ T cells in the paired patient analysis in response to anti-TNF therapy. Each differential region was assigned to two genes, which totalled to 232 potentially regulated genes. This list of genes was used to intersect with the differentially expressed genes in the paired RNA-seq analysis, of which 288 genes were identified. 10 genes were found overlapping between the two analyses, which are listed in Table 4.11, (P= 0.02, Fishers test).

PTGER4 is differentially regulated in $CD8^+$ T cells in treated AS patients

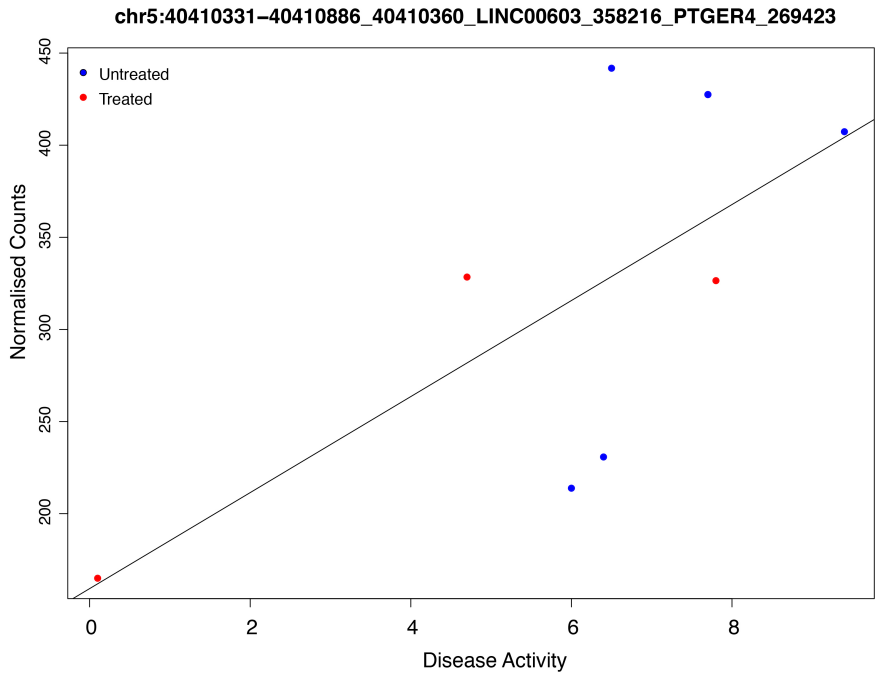
One of the overlapping genes was the top differential open chromatin region that lies close to *PTGER4*. In both analyses, there was a significant down-regulation of *PTGER4* gene expression or open chromatin at chr5:40410331-40410886 in the AS patients treated with TNF-α inhibition therapy. This gene was reported in AS GWAS (Cortes et al. 2013; Evans et al. 2011), and SNPs in *PTGER4* are known to affect severity of AS (Ozen et al. 2017; Chai et al. 2013). When plotting the normalised ATAC-seq and RNA-seq counts, the expression levels demonstrate a correlation with disease activity in both (Figure 4.13). The protein encoded by this gene is one of four G protein-coupled transmembrane receptors involved in prostaglandin E2 (PGE2) signalling, regulated by cyclooxygenase (COX) activity, the target for NSAIDs. A UCSC genome browser snapshot demonstrating this peak is shown in Figure 4.14. The peak also overlaps two GWAS SNPs for IBD, CD and UC (rs11742570 and rs6451493) (Lange et al. 2017; Liu et al. 2015; Julia et al. 2013; Jostins et al. 2012; Anderson et al. 2011; Franke et al. 2010). FOS and JUN are shown to bind in this location, and both are also down-regulated in the TNF-α treated individuals. There are also many strong promoter capture Hi-C

interactions around this region interacting with the *PTGER4* promoter, but these do not directly fall on this peak.

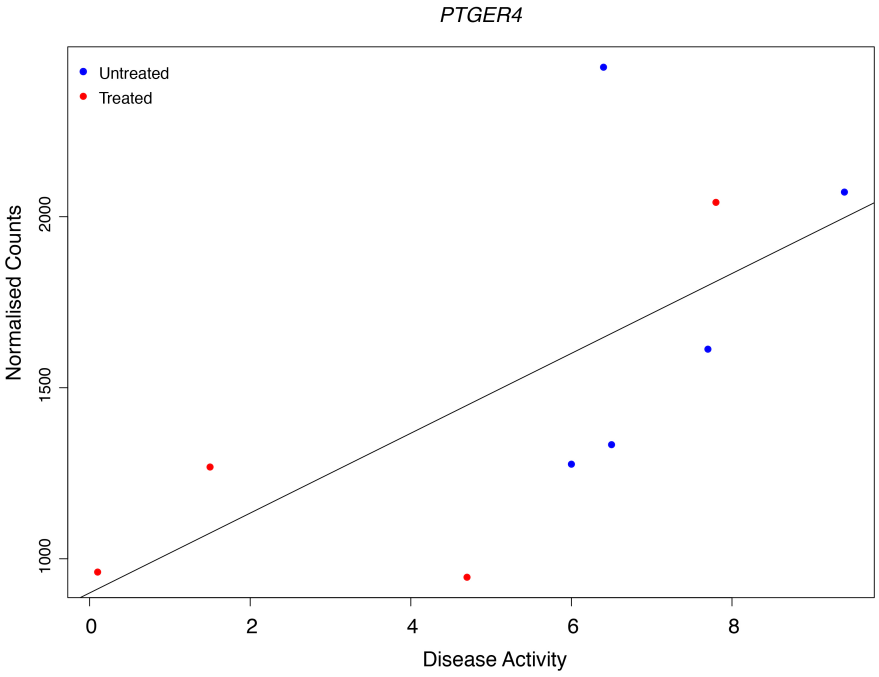
Other genes that may be differentially regulated are *JUN* and *FOS* (these dimerise to form the TF complex AP-1) involved in TGF-β signalling. They have roles in the development of cells involved in the maintenance of the skeleton, through signal transduction, cell proliferation and differentiation. *CYLD* codes for a deubiquitinase, involved in the regulation of NF-κB regulation by inhibiting its nuclear translocation. *PFKFB3* is required for the prevention of apotosis, and blocking its activity has demonstrated some therapeutic effect (Schoors et al. 2014). HIVEP1 can bind to the promoter of MHC class I genes (Veer et al. 1992), along with *IL2R* and IL-β related genes, and is known to act in T cell activation.

CD8$^+$	**RNA-seq DE in treatment group**
PTGER4	down-regulated
JUN	down-regulated
ADTRP	down-regulated
CYLD	down-regulated
YPEL5	down-regulated
OXNAD1	down-regulated
PFKFB3	down-regulated
PNP	up-regulated
HIVEP1	down-regulated
FOS	down-regulated

Table 4.11: Differentially expressed genes in anti-TNF therapy samples compared with naive AS patients, which are located proximal to a differentially open chromatin region in CD8$^+$ T cells. All ATAC-seq differentially open chromatin regions were found more closed or "down-regulated" in the treatment group. All RNA-seq genes apart from one were also found down-regulated. Genes included passed an FDR <0.05.



(a)



(b)

Figure 4.13: Scatterplot of normalised counts for *PTGER4* expression in CD8$^+$ T cells in relation to disease activity. **(a)** Treated versus untreated AS patients at the differentially open peak close to *PTGER4* (r^2=0.48, *P*=0.057) and **(b)** Treated versus untreated AS patients *PTGER4* mRNA counts (r^2=0.45, *P*=0.048), both showing increased expression is related to increased disease activity. AS patients treated with TNF-α inhibition therapy mostly demonstrate lower expression levels, and the patients with the lower BASDAI score demonstrated the lower read count in ATAC-seq.

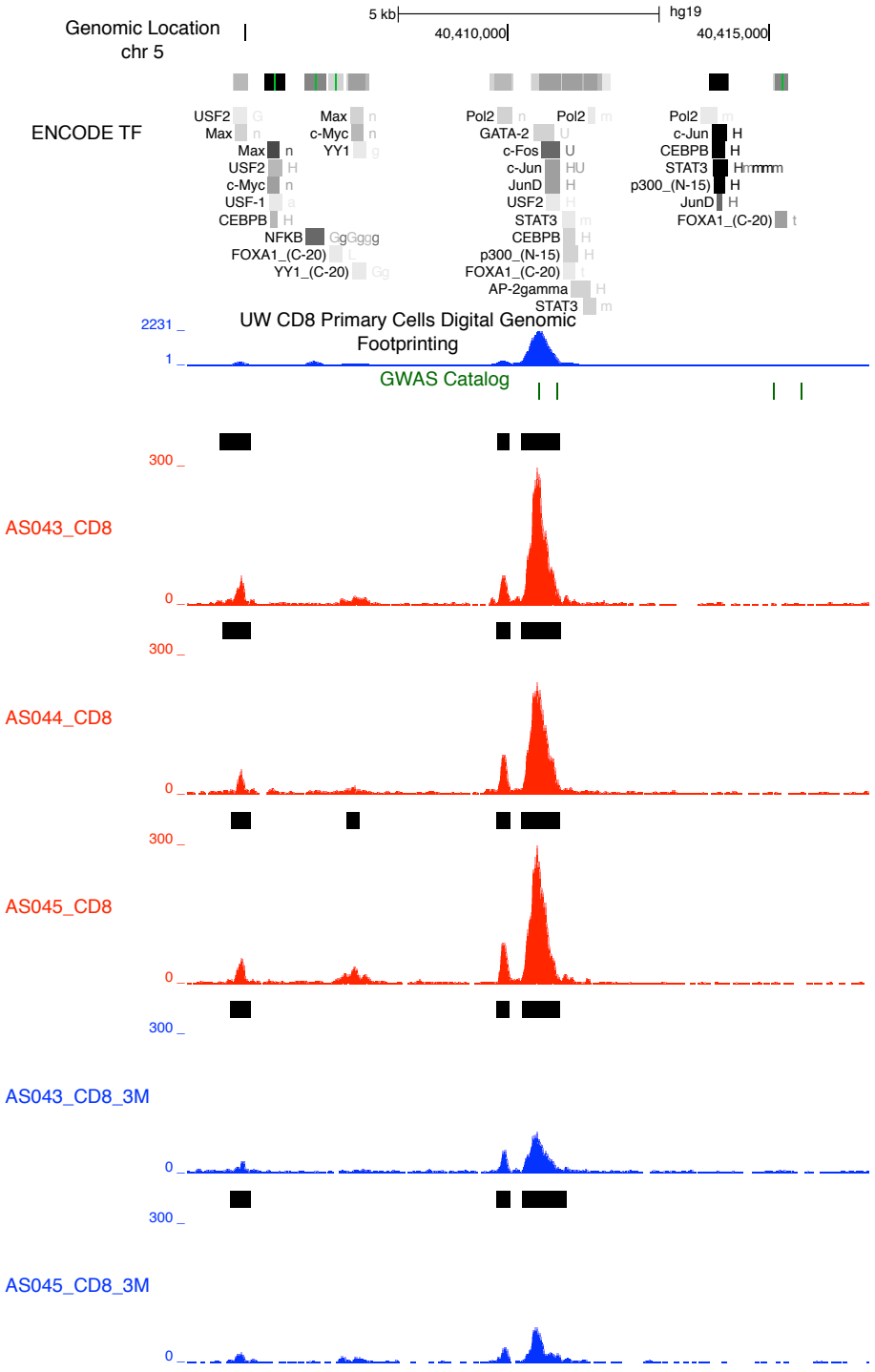


Figure 4.14: Differentially open ATAC-seq peak in CD8$^+$ T cells, demonstrating a down-regulation in patients with 3 months TNF-α inhibition. UCSC Genome-browser view showing the ATAC-seq read density distribution of 3 AS patients and 2 of their paired 3 month samples after anti-TNF therapy. The peak also overlaps a GWAS associated SNP for IBD and a DNase-I digital footprinting site in primary CD8$^+$ T cells, along with the binding of multiple transcription factors including JUN and FOS.

4.4 Discussion

The results in this chapter have identified differential open chromatin regions between AS patients and healthy controls, by successfully defining the open chromatin landscape in AS patients in 3 immune cell types. Differentially open regions between AS patients and their paired 3 month anti-TNF therapy follow up sample were also described in $CD8^+$ T cells, along with the transcriptional landscape in AS patients treated with TNF-α inhibitors ETA and ADA in 3 cell types. The possible differentially expressed genes that may be linked to changes in chromatin accessibility in these comparisons were also investigated, with the aim to identify possible mechanisms of gene and pathway dysregulation.

4.4.1 Differential chromatin accessibility in AS patients

This chapter describes a relatively new approach to define differences between disease and healthy individuals, by analysing cell specific chromatin accessibility data in primary immune cells types, isolated from blood collected on the same day from the donor. This approach is believed to have preserved the integrity of the epigenetic landscape; providing information regarding the accessibility of chromatin regions relating to disease state. The analysis described is an example of one way that chromatin accessibility can be used to define differences between two states in the same cell type. A merged master list of peaks present in two or more individuals will remove peaks that appear in only one sample, and therefore likely to be spurious. This approach will not, however, account for the absence of a peak in one set of samples, as only peaks that appear in both are taken forward for analysis. One similar study was performed in 3 SLE patient samples that had been biobanked for two years along with one freshly obtained SLE sample, and 4 healthy controls as a comparison in $CD19^+$ B cells (Scharer et al. 2016). This study was able to detect changes in differentially accessible peaks

which were associated with leukocyte differentiation, cellular activation, and B cell activation.

Main findings following integration of differential open chromatin and differential gene expression

Differential analysis between AS patients and healthy controls identified a large number of regions in $CD14^+$ monocytes, with a smaller number for $CD4^+$ and $CD8^+$ T cells. Innate immune cell types may be more receptive to change by an inflammatory stimulus, and may be more likely to show increased changes in chromatin accessibility globally. Pathway enrichment analysis for $CD14^+$ monocyte ATAC-seq and RNA-seq integration found enrichment in many genes involved in NF-κB signalling and genes up-regulated during the unfolded protein response, reproducing some pathways in the main RNA-seq analysis (Chapter 3). *NFKBIA* gene expression was consistently down-regulated in AS patients, and also demonstrated a down-regulated chromatin region in AS patients (data not shown). The differential open chromatin regions at the *IRF1* gene was highly significant in $CD14^+$ monocytes, but did not correlate with gene expression levels, requiring further investigation. These differential regions lie in proximity to 2 GWAS SNPs, rs4705952 is associated with CRP levels and chronic inflammation (Dehghan et al. 2011), and rs2549003 which is an asthma risk allele (Myers et al. 2014). The decrease in chromatin accessibility suggests that *IRF1* is transcriptionally down-regulated in AS patients, and a meta-analysis in human monocyte-derived macrophages revealed that *IRF1* was negatively regulated in SpA individuals (Fert et al. 2014). The $CD4^+$ T cell differential regions only demonstrated one overlap with RNA-seq data, located proximal to the gene *RGS1*. *RGS1* was found down-regulated in both the ATAC-seq and RNA-seq analysis in both T cell subtypes, suggesting that it may play a potentially important gene in AS T cell responses. RGS1 desensitises the chemokine

receptors CCR7 and CXCR4 that are critical to the localization of T and B cells in lymphoid organs (Caballero-Franco and Kissler 2016). *RGS1* was also found reduced in abundance in SLE patients (Bradley et al. 2015), and is associated with several other autoimmune diseases such as MS, celiac disease and T1D (Johnson et al. 2010; IMSGC 2010; Hunt et al. 2008; Smyth et al. 2008). *RGS1* has been highlighted as a biomarker for undifferentiated spondyloarthritis, where *RGS1* was inducibly higher in undifferentiated SpA than classified AS patients (Gu et al. 2009). RGS proteins are GTPase-activating proteins that modulate chemokine receptor signalling and *RGS1* is highly expressed in lymphoid organs serving as a negative regulator of chemokine receptor signalling in lymphocytes (Moratz et al. 2004). A recent study identified a link between *RGS1* expression and T follicular helper (TFH) cell frequency, providing a possible explanation for the association of *RGS1* with several autoimmune disorders (Caballero-Franco and Kissler 2016). Therefore understanding whether *RGS1* may be affecting AS pathogenesis through the regulation of TFH cells may be an interesting candidate for future research.

The majority of differentially open chromatin regions in $CD14^+$ monocytes are down-regulated, suggesting a general effect across the genome of gene expression down-regulation. This may relate to the significantly increased number of HDACs described in $CD14^+$ monocytes in Chapter 3, causing many TF sites to become deactivated. It was shown that HDAC inhibitors are strongly associated with a gain in chromatin accessibility, further reinforcing a future role of HDAC inhibitors as a treatment option, and could be investigated by treating $CD14^+$ monocytes with a HDAC inhibitor and assessing global chromatin accessibility (Qu et al. 2017).

The current limitations of this approach

The inability to find significantly enriched differentially open chromatin regions in $CD4^+$ and $CD8^+$ T cells may relate to a number of limitations in this study. Firstly the sample size is small for this type of genome-wide analysis, so we may currently lack the power to detect meaningful regions. The current approach of only analysing peaks identified in a number of samples may be too stringent, overlooking a large amount of variation that may be particular to an individual or a set of individuals. This approach also relies on the parameters set by the peak caller to define open chromatin regions. A strategy that removes the requirement for peak calling may be interesting to investigate. One such software is called diffReps (Shen et al. 2013), that is able to count the reads in a pre-defined sliding window and outputs a list of differential sites. This approach will also be able to detect regions that are completely down-regulated and therefore lack a peak in one set of samples that otherwise would have been unaccounted for. Examples demonstrating this are discussed in Chapter 5, Section 5.3.5. However, most allelic differences in expression were reported not to be on/off events, but instead reflect biases in the level of expression from each allele (Dixon et al. 2015).

4.4.2 The effects of TNF-α inhibition

Genome-wide profiling on the effect of TNF-α inhibition in AS patients is still relatively unknown, with only two previous studies investigating the effects of treatment on gene expression (Wang et al. 2017; Dolcino et al. 2017). Dolcino et al found genes are differently modulated in patients that show a good response to treatment compared to non-responders, and that ADA treatment normalised the altered expression of genes involved in AS-associated pathways. Wang et al carried out a similar approach to our study by comparing AS patients pre- and post- treatment, identifying many differentially expressed genes in

PBMCs. A small number of these overlapped with GWAS reported genes. Our patients were treated with ADA and ETA, and were all described as patient responders. The previous studies identified genes in treatment response in signalling pathways involving MAPK, fibroblast growth factor receptor and toll-like receptor signalling and cytokine-cytokine receptor interactions. This analysis demonstrated some overlap with these results, with damping of genes involved in the TNF and NF-κB signalling pathways, increased expression of *MAPK14* and many pathways involving cytokine receptors. However, it may not be ideal to compare these results, since studies conducted in PBMC populations may actually be reflecting the changes in the proportions of PBMCs after treatment. The upregulation of *CASP8* in both $CD14^+$ monocytes and $CD8^+$ T cells may point to an important role of apoptosis in treatment response. CASP8 is also involved in the negative regulation of necroptosis, and decreased CASP8 has been shown to trigger inflammation in vivo by sensitizing cells to RIPK3-mediated necroptosis (Oberst et al. 2011). Deletion of *CASP8* in mouse macrophages was shown to promote onset of a mild systemic inflammatory disease (Cuda et al. 2015) and may dictate the response to TLR activation. There is a growing body of evidence suggesting the importance of cross talk between innate immune inflammatory signalling and apoptosis, associated with a large number of autoinflammatory diseases (Feltham et al. 2017). The gene expression of CASP8 is increased once TNF-α inhibition has occurred, and may act to restrict lymphocyte accumulation and therefore decrease the inflammatory response (Su and Lenardo 2008).

Differentially regulated lncRNAs following anti-TNF treatment

This study was able to introduce further novelty by analysing lncRNAs that are differentially expressed after treatment. *NEAT1* was found to be significantly down-regulated after treatment, and was consistent with the up-regulation of

NEAT1 in AS patients from Chapter 3. *NEAT1* levels were found to be related to disease activity in AS and SLE (Zhang et al. 2016). Zhang et al found that *NEAT1* may dysregulate TLR4 signalling and chemokine activity and may induce late stage phosphorylation of the MAPK signalling pathways (Zhang et al. 2016). It may therefore act as a candidate biomarker of treatment response and disease severity in both AS and SLE patients, and is an interesting candidate for future studies. Another interesting lncRNA is *HOTAIRM1*, which was found down-regulated in treated AS patients. Knockdown of *HOTAIRM1* was shown to reduce myeloid differentiation associated genes, and therefore may play a role in regulating normal myelopoiesis by regulating *HOXA* gene expression (Bei et al. 2007). Further understanding the role of lncRNA in regulating gene expression in regards to the development of immune cells following treatment may be an important area in future studies. *SCARNA2* was found up-regulated in the treatment group, and was also found over-expressed in IBD patients treated with azathioprine (DMARD) compared to pre-treatment (Coulthard et al. 2017). *SNHG15*, down-regulated in $CD14^+$ monocytes and $CD8^+$ T cells, has been demonstrated to be involved in cell proliferation and invasion in cancer (Chen et al. 2016) and also in response to chemical stressors (Tani et al. 2014). Polymorphisms in *HCP5* (HLA complex P5), found up-regulated in $CD14^+$ monocytes and $CD8^+$ T cells, is located in the HLA region. It has previously been associated with psoriasis and SLE (Li et al. 2014b; Ciccacci et al. 2014). Carriers of *HCP5* variants demonstrate a reduced viral load, which suggests they may mount a strong immune reaction leading to excessive inflammation.

4.4.3 Chromatin accessibility changes in response to anti-TNF therapy

PTGER4 is an interesting gene that was found differentially expressed in all 3 cell types post-treatment. A differential open chromatin region close to *PTGER4*

was found differentially closed post-treatment, correlating with the reduced expression of this gene in $CD8^+$ T cells. Many promoter capture Hi-C contacts were found surrounding this region, and although the interaction was not found directly at the differential peak, this technique lacks resolution due to the large interacting regions it creates. A higher resolution technique such as Capture-C (Hughes et al. 2014) may be more accurate at detecting enhancer regions that interact with promoters. PGE2 is known to promote inflammation and skeletal related changes in RA (Robinson et al. 1975; Lippiello et al. 1978), and reducing PGE2 levels can reduce inflammation in a mouse model of poly-arthritis (Portanova et al. 1996). It has been demonstrated that EP4 (the receptor encoded by *PTGER4*) deficient mice demonstrate a similar phenotype of reduced PGE2 levels, reducing the severity and incidence of the clinical signs of arthritis (McCoy et al. 2002). Signalling through EP4 can induce T cell activation, and the reduced expression of the EP4 receptor in the treatment group may cause decreased T cells responses. Patients treated early with TNF-α inhibitor therapy demonstrate reduced radiographic progression (Maksymowych et al. 2013), suggesting there may be a mechanisms that prevents new bone formation in these patients. *PTGER4* expression in $CD8^+$ T cells may act as a biomarker to measure disease activity and treatment response, and should be prioritised for future follow up studies. The transcription factors *JUN* and *FOS* were also found down-regulated in $CD8^+$ T cells post-treatment response. These TFs regulate TGF-β signalling, regulating the development of cells destined to form and maintain the skeleton. These TFs were also found to bind around the *PTGER4* differentially open chromatin region and may therefore be involved in regulating *PTGER4* expression.

4.4.4 Prioritised targets for future validation

A number of findings should be taken forward and validated in a larger AS cohort, or investigated using functional assays. The down-regulation of *RGS1* levels and the differentially expressed peak in AS T cells should be validated in a larger AS cohort due to its known importance in T cell signalling. The down-regulation of GWAS locus *PTGER4* in treated AS patients $CD8^+$ T cells should be investigated further. This differentially open chromatin peak sits on a SNP found associated with IBD, a disease closely related to AS. A future hypothesis may be to understand whether there is a genotypic effect of this SNP on chromatin accessibility and TF binding using ChIP-seq, and its influence on *PTGER4* levels. This is the first study to investigate lncRNAs in AS in a cell type specific manner. The identification of lncRNA *NEAT1* in both $CD14^+$ monocyte comparisons and its correlation with disease activity in AS and other autoimmune diseases may suggest its involvement in disease aetiology. Further studies should investigate the functional mechanisms of *NEAT1* on target genes differentially expressed in AS.

4.4.5 Limitations

Low sample sizes and the lack of a gold standard for bioinformatic analysis in ATAC-seq are current limitations that may explain the inability to detect significant changes in chromatin accessibility in some comparisons. Further methods of analysis require investigation, along with continued recruitment of patients through the GENExpress ID study. Since AS is known to be clinically heterogenous, the number of samples required is difficult to predict. Using the new optimised method for ATAC-seq, called FAST-ATAC (Corces et al. 2016) may also reduce some of the noise found in our data, increasing the sensitivity to detect changes and also reducing the number of mitochondrial reads to improve the number of filtered reads after sequencing. ChIP-seq of AS patients would

add to the current level of data, which can be integrated to further refine disease relevant chromatin states by ChromHMM.

4.4.6 Conclusion

Chromatin profiling may be an important tool in studying immune-mediated disease related changes. The chromatin landscape can begin to reveal current but also future potential gene expression changes in response to an inflammatory stimulus. RNA-seq data can provide a large amount of information regarding pathways related to disease, but may not provide a view on the underlying mechanisms promoting the change in gene expression. This pilot study has been able to demonstrate the cell type specific enhancer differences relating to changes in differentially expressed genes. The prioritisation of AS specific enhancers will promote future research in identifying the molecular mechanisms that may be involved in changes in the regulatory landscape in AS patients and other inflammatory diseases, providing the possibility for a stratified approach in treating related immune-mediated diseases.

Chapter 5

Prioritisation and interpretation of AS GWAS SNPs

5.1 Introduction

5.1.1 AS GWAS and the IGAS UK cohort

Genome wide association studies have made a fundamental contribution in the identification of AS susceptibility loci (Burton 2007; Reveille 2010; Evans et al. 2011; Cortes et al. 2013). The most important AS association study to date based on sample size and coverage in immune revelant loci has been the Immunochip study, involving 10,619 cases and 15,145 controls from European and Asian ancestry (Cortes et al. 2013). The Immunochip study includes high-density coverage of the top 2,000 independent association signals from 11 distinct autoimmune and inflammatory diseases (Trynka et al. 2011; Cortes and Brown 2011), interrogating 195,806 SNPs and 718 small insertion-deletions. This study identified 13 additional susceptibility loci, along with confirming a number of associations from previous studies, totalling 43 loci (Cortes et al. 2013). A more recent cross-disease meta analysis was able to identify further AS susceptibility loci, totalling 113 distinct association signals (Ellinghaus et al. 2016).

The specific purpose of the Immunochip study was deep replication of currently known GWAS susceptibility loci, and also refinenement of these signals by fine-

mapping due to the increased SNP coverage at immune relevant loci. The Immunochip contains 186 immune related loci that had been chosen based on current knowledge of their relevance in immune mediated diseases, with dense SNP coverage. There are a number of limitations to the Immunochip study, leaving a large proportion of the susceptibility loci unable to be extensively fine-mapped. A study prioritising AS SNPs only demonstrated a 10% narrowing in 4 of 8 loci of associated regions; the resolution being limited due to extensive LD (Trynka et al. 2011). Fine mapping using the Immunochip has demonstrated success in some diseases, such as coeliac disease, in which a single gene was assigned to 29 of the 54 non-MHC loci, containing a credible set of SNPs for each (Trynka et al. 2011).

5.1.2 GWAS and its limitation in causal variant identification

As previously discussed, GWAS often fail to correctly identify the causal variant due to LD (Visscher and Hill 2009). In some cases, the lead SNP may be in high or complete LD with hundreds of SNPs, creating difficulty in refining the signal at the locus. The Immunochip study was unable to successfully fine-map SNPs in the majority of AS susceptibility loci, therefore further refinement using statistical and functional data is required to narrow down likely causal SNPs and genes in AS pathogenesis. It is also important to note that low frequency variants (MAF <0.5%) are often not included in GWAS arrays, but are suspected to account for large effect sizes and may explain the missing heritability estimates that are found in many common diseases (Manolio et al. 2009). There are a number of limitations specific to the Immunochip, which was designed using early 1000 genome pilot data (February 2010 release), meaning that the coverage of low-frequency and rare variant spectrum is incomplete. It may also be the case that a causal variant may not have been genotyped due to the Immunochip's focus on specific immune related loci, possibly missing areas of the genome of

unknown importance. The Immunochip also does not type rare CNVs and the design is also based on the haplotype structure and SNPs identified in Europeans, therefore non-European variation may be under-represented (Cortes and Brown 2011). See Chapter 1, Section 1.6 for more details.

5.1.3 Methods used to fine map candidate SNPs

Aside from the use of custom genotyping arrays such as the Immunochip, and more recent arrays that enhance the representation of low-frequency and rare variants such as The UK Biobank Axiom Array (http://www.ukbiobank.ac.uk), there are other methods that increase coverage and access to low-frequency and rare variants. One such way is genotype imputation, which is a cost-effective strategy to expand the SNP content of genome-wide genotyping arrays. This relies on reference panels of SNPs, generated by the HapMap project or more recently the 1000 Genomes project (pilot, phase I and phase III). Phase III included a total of 88 million variants which were phased onto high-quality haplotypes, from the genomes of 2504 individuals from 26 populations and include >99% of SNPs with a MAF of >1% (Auton et al. 2015). More recent imputation panels from the UK10K project have identified 42 million SNVs and 3.5 million INDELs, of which 24 million were novel variants (Walter et al. 2015). The use of 1000 Genomes and the UK10K combined may increase the ability to discover low-frequency variants associated with disease (Zheng et al. 2015), and to improve the ability to fine-map SNPs associated with disease. The various fine-mapping methods are described in more detail in Chapter 1, Section 1.6.6.

5.1.4 Genomic annotation using publically available datasets

Often these fine mapping strategies are not able to refine the signal to a single variant, therefore interpretation with genomic annotation is often necessary to build a complete picture of the functional relevance of the

SNPs identified across different cell types. This includes epigenetic (DNA methylation), epigenomic (chromatin accessibility and histone modifications) and transcriptomic annotations, along with chromosome conformation data (promoter capture Hi-C, Capture-C). These annotations are all interconnected and when integrated may give us further confidence in how a variant may be regulating gene expression. They are also highly cell type specific, and may vary with different environmental stimuli, therefore there has been a major drive in generating annotations in a variety of primary cell types in recent years through various consortium projects (Dunham 2012; Adams et al. 2012; Bernstein et al. 2010). One way to integrate some types of this data is through a software tool called chromHMM, developed by Ernst et al (Ernst and Kellis 2012). More details on these datasets and methods for integration are mentioned in Chapter 1, Section 1.8.4. Fine mapping results from GWAS can be re-weighted by integrating these various functional datasets using tools such a FGWAS (Pickrell 2014).

5.1.5 Chromatin Conformation Capture to identify enhancer-promoter interactions

Genes regulated by cell specific enhancers are often located at a distance from the disease associated SNP and therefore may be difficult to define, especially when disease aetiology is not well understood. One experimental technique, chromatin conformation capture, is able to link enhancers to their target gene promoters. There have been a number of developments in the conformation capture method, including 3C, 4C, 5C, Hi-C and Capture-C, which vary in their ability to define a number of regulatory regions (Fraser et al. 2015). For 3C, a candidate promoter and enhancer is already required, and PCR primers are designed at intervals between them to test for an increase in signal at the promoter and enhancer using RT-qPCR (Hagege et al. 2007). Hi-C and Capture-

C use high-throughput sequencing to sequence all regions that are found linked to the candidate region that is enriched using biotinylated probes (Mifsud et al. 2015; Hughes et al. 2014). Recently, Hi-C was performed at gene promoters across a number of different cell types and activation states, providing a catalogue for chromatin interactions across different cell types (Javierre et al. 2016). There is also a web interface for easy interrogation of a SNP or region of interest (https://www.chicp.org) (Schofield et al. 2016).

5.1.6 Functional characterisation of GWAS SNPs

Once a SNP has been prioritised to be potentially causal, functional follow up studies are required to define the causality of a variant, and mechanism of action. Reporter gene assays using traditional cloning techniques have been a popular method in identifying enhancer elements (Dailey 2015; White 2015), and more recently the CRISPR-cas9 system has been widely adopted in the attempt to dissect the *cis*-acting effects of genetic risk variants. Genome editing of human pluripotent stem cells can allow differentiation into primary cells of interest to allow a more accurate assessment of the enhancer or variant in the relevant cell type (Sanjana et al. 2016; Diao et al. 2016; Korkmaz et al. 2016; Xue et al. 2016). CRISPR/cas9 editing of a single base pair in the genome has also been achieved using programmable site-specific nucleases, allowing for more concrete confirmation of the causality between SNPs and disease, and has potential value in clinical applications (Govindan and Ramalingam 2016).

5.2 Aims

This chapter aims to fine-map disease association signals in AS and annotate with differentially open chromatin and expressed genes in AS. This will provide evidence for putative functional variants and modulated genes. The specific aims of this chapter are to:

1. **Perform association and conditional analyses on the UK IGAS cohort.** Conditional analysis will highlight any additional association signals at each locus.

2. **Perform fine mapping for each distinct association signal by generating credible sets of SNPs and their posterior probabilities.** For each locus, fine-mapping will identify a credible set of SNPs that best explain the association at that locus, assuming there is one causal SNP at each association signal.

3. **Integrate fine-mapped SNPs with regions of differentially open chromatin identified through ATAC-seq, and GWAS reported genes with RNA-seq differentially expressed genes.** The integration of data may provide further evidence of potential causal mechanisms acting through GWAS associated variants.

4. **Give an example of a functional experiment that can be performed with fine-mapped SNPs.** Using a variety of functional assays, an AS associated *CARD9* locus SNP was tested for potential function relevance in altering gene expression.

5. **Hypothesise the functional mechanism of fine-mapped SNPs based on interaction with cell specific chromatin accessibility data, chromatin interaction data and gene expression.** The causal mechanism of a fine-mapped SNP at locus 2p15 is hypothesised to be functional in a cell type specific manner by the integration of various data sets.

5.3 Results

5.3.1 Fine mapping of AS associated loci

Comprehensive fine mapping was carried out in 26 of the reported non-MHC susceptibility loci from the IGAS Immunochip study. I was given access to a subset of the European cohort of AS patients, more specifically the UK cohort, which includes 4101 AS cases and 9700 controls. Phasing and imputation was carried out in a 2Mb area surrounding the lead SNP, using the 1000 Genome Project Phase 3 Version 3 reference panel (October 2015 release). This provides almost complete coverage of common and low frequency variation across these loci. Well imputed SNPs (found in at least 70% of the sample size) were carried forward for association analysis. Filtering these SNPs is important in the credible set generation, so that the results are not dependent on the quality of the imputation. The posterior probability is determined by effect size, which can be directly affected by the quality of the imputation. Credible sets of SNPs were constructed containing SNPs most likely to be driving the association signal, along with their posterior probability. The 50% and 95% credible set was generated for each signal, and SNPs were ranked according to their approximate Bayes factor, including further variants until their cumulative posterior probabilities exceeded 0.99 (Bunt et al. 2015). Methods are described in more detail in Chapter 2. Fine mapping analysis in a subset of the Immunochip AS cohort was beneficial for future prioritisation, as it provided a good refinement in a large number of loci as the sample size was still substantially large.

5.3.2 Defining variants driving AS association signals

At each association signal genome wide significant associations were identified, characterised by a log10 approximate Bayes factor of 5 or above, as shown in Table 5.1. Each signal is described by the lead SNP that explains the largest part of the association signal, along with the log10 approximate Bayes factor and the number of SNPs contained in the 95% or 50% credible set, each with their own posterior probability. In most cases, the SNPs identified in the UK IGAS cohort differed from the European analysis, demonstrating the limitations of this method, which may depend on the power you have to detect associations along with the set of individuals in the cohort. The success of fine-mapping is correlated with the effect size and relative allele frequency (RAF) at each locus (Bunt et al. 2015). Improved identification of causal SNPs may arise when overlapping the credible set with epigenomic data such as open chromatin and enhancer or promoter methylation marks.

Approximate bayesian fine-mapping of these 26 loci results in median 50% and 95% credible sets containing 10 (1-2952) and 25 (3-10473) variants respectively. 10 of the 26 loci were reduced to fewer than 10 SNPs in the 95% credible set. However, only 19 of the 26 credible sets included the lead variant reported in the lead GWAS analysis (Cortes et al. 2013). The analysis with the European cohort reported 24 loci containing the lead GWAS SNP, clearly showing that successful fine-mapping is sample size dependent, and therefore its utility depends on the availability of large-scale genotyping data (Bunt et al. 2015). The resolution is most precise at loci with strong association signals, therefore the credible sets generated from *IL23R*, *KIF12B*, locus 2p15, *GPR35*, *ERAP1*, and locus 21q22 can be assumed to be relatively accurate. Some of these association plots are shown in Figure 5.1. For majority of the credible sets with less than optimal resolution, integration with epigenomic data is still able to make use of this data.

The greatest refinement was achieved for *ERAP1* with the primary signal reaching a log10 approximate bayes factor of 23.38. This gene is already known to be important in AS pathogenesis and therefore is still present at a very high significance in the smaller UK cohort. The intergenic region on chromosome 2p15 was the second most significant association, with a log10 approximate bayes factor of 18.42. This signal was refined to a 95% credible set of 3 SNPs, therefore the probable causal SNP can easily be identified with epigenomic data (Figure 5.8a). Those loci that were unable to be fine-mapped using this cohort are listed in the Appendix (Chapter 7, Table 7.7). The index variant rs1128905 for *CARD9* was found to be a genotyping error and therefore is not present in the credible set (7, Figure 7.4). However, the index variant from a previous GWAS rs10781500 is present (Evans et al. 2011). In *ERAP1* the index variant was not present, but SNPs with an r^2 of 1 with the index SNP are found in the credible set.

Table 5.1: Fine mapping of Immunochip AS associated loci. 17 AS associated loci were fine mapped to a 50% and 95% credible set, *8 of which either did not reach significance, or compared with the analysis in the European cohort were different. Also included are the results from the European *CARD9* locus analysis. Any additional signals identified through conditional analysis are also shown. MAF= minor allele frequency, OR= odds ratio, ABF= approximate bayes factor, PP= posterior probability.

Locus	MAF	OR	ABF variant	log10 ABF	PP	50% & 95% credible set	Index SNP present
1p31 (*IL23R*)	0.03	1.69	1:67699535:GTT:G	10.9	0.15	6 & 15	No
-	0.37	1.21	rs12141575	6.04	0.09	7 & 13	Yes
1p26 (*RUNX3*)	0.46	1.18	rs6600247	4.23	0.85	1 & 1322	Yes*
1q21 (*IL6R*)	0.4	1.16	rs12126142	13.5	0.75	1 & 3	Yes*
1q23 (*FCGR2A*)	0.23	1.13	rs2165088	2.51	0.03	2952 & 10473	Yes*
1q32 (*KIF12B*)	0.25	1.22	rs7522462	8.73	0.1	7 & 20	Yes
2p15	0.42	1.33	rs4672505	18.42	0.43	2 & 3	Yes
2q12 (*IL1R1*)	0.35	1.17	rs4851529	5.66	0.85	4 & 10	Yes
2q37 (*GPR35*)	0.2	1.21	rs34236350	6	0.62	1 & 3	Yes
5p13 (*IL7R*)	0.34	1.15	rs2172754	5	0.42	2 & 40	No*
5q15 (*ERAP1*)	0.40	1.33	rs26510	23.38	0.13	5 & 12	No
-	0.45	1.05	rs2910686	11.6	0.13	25 & 77	Yes
-	0.18	1.33	rs13178387	5.38	0.22	9 & 24	Yes

5q33 (*IL12B*)	0.29	1.15	rs6897260	4.62	0.21	3 & 7	Yes*
9q32 (*CARD9*) UK	0.45	1.14	rs4567159	5.7	0.48	2 & 5	No*
9q32 (*CARD9*) EUR	0.42	1.10	rs35670379	6.16	0.1	6 & 26	Yes
10q22 (*ZMIZ1*)	0.43	1.16	rs1250555	5.35	0.48	2 & 9	Yes
10q24 (*NKX2-3*)	0.26	1.19	rs11190133	6.26	0.39	6 & 46	Yes
12p13 (*TNFRSF1A*)	0.34	1.15	rs7954567	5.27	0.95	1 & 1	Yes
-	0.52	1.14	rs4149577	4	0.66	1 & 6	Yes
14q31 (*GPR65*)	0.1	1.2	rs3742704	4.97	0.2	3 & 9	Yes
17q11 (*NOS2*)	0.2	1.2	rs28944192	5.16	0.28	4 & 27	No*
17q21 (*TBKBP1*)	0.46	1.13	rs4794057	4.24	0.02	58 & 191	Yes*
21q22	0.21	1.26	rs9977672	10.7	0.2	4 & 10	Yes

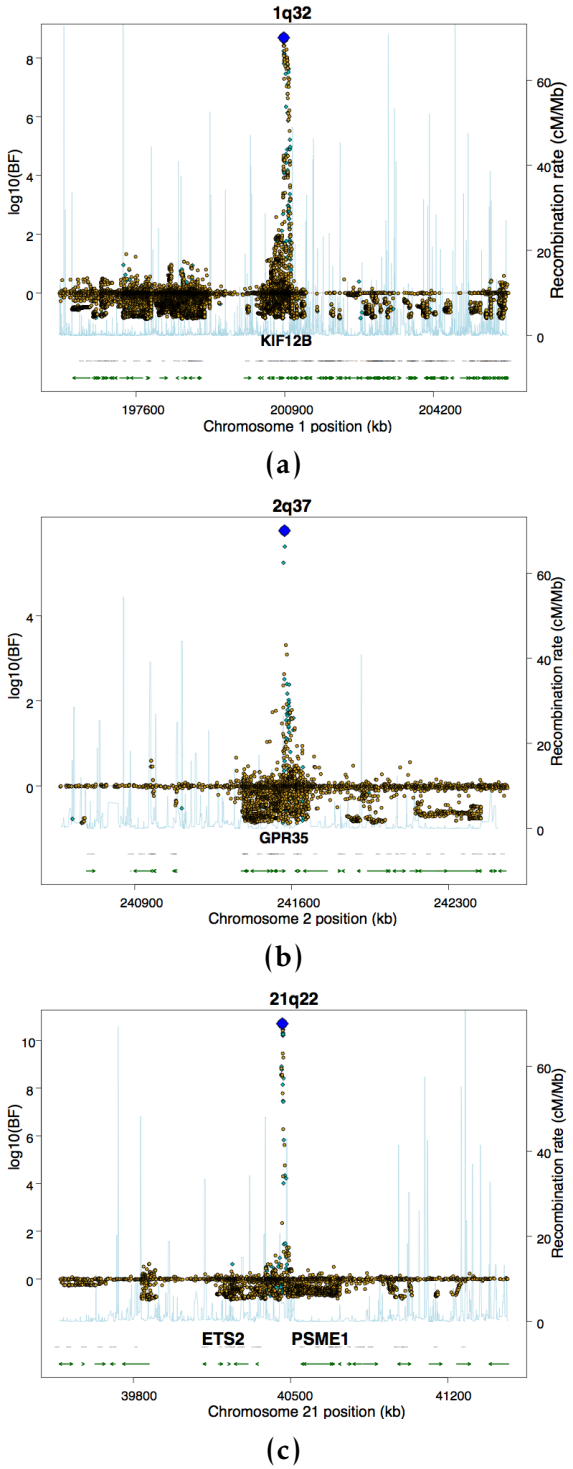


Figure 5.1: Local association plots of the log10 approximate bayes factor for the most significant loci. at **(a)** 1q32 *KIF12B* **(b)** 2q37 *GPR35* **(c)** gene desert locus 21q22. Recombination hotspots are also demonstrated in blue, and imputed SNPs are shown in gold.

5.3.3 Independent association signals at AS susceptibility loci

Understanding the genetic architecture at associated loci can begin by describing any distinct association signals, which may arise due to multiple independent causal variants. In this analysis, this was achieved by conditional analysis using SNPTEST with a Bayesian additive model (Marchini et al. 2007; Marchini and Howie 2010). Independent association signals were found for three of the loci; *IL23R*, *ERAP1* and *TNFRSF1A*, as shown in Table 5.1. These loci were also reported to have secondary signals in the original Immunochip paper (Cortes et al. 2013). The *ERAP1* locus has extensive LD, meaning causal SNP identification has proven difficult. Evans et al proposed a two mutation model, with the primary effect due to the locus containing rs30187, and a secondary effect by one of the other two signals. rs30187 did not appear in our credible set, but SNPs in complete LD were present. Associations in the *IL23R* are shared between many inflammatory diseases, including psoriasis, psoariatic arthritis and Crohn's disease (Liu et al. 2015; Tsoi et al. 2012). The AS Immunochip paper found two independent signals associated with the *IL23R*. The primary signal identified by fine mapping did not contain the main Immunochip SNP, but instead the top variant was an INDEL (67699535:GTT:G). This INDEL was also shared with a bayesian fine mapping approach within the group for psoriatic arthritis (unpublished). The primary and secondary associations at the *TNFRSF1A* locus were found to be reversed when fine mapping, with rs7954567 appearing alone as the primary signal, and the original primary SNP appearing within the credible set of six SNPs as the secondary signal. SNPs in LD have been described to affect splicing and transcriptional control of *TNFRSF1A* (Rittore et al. 2014). These loci are also shared with inflammatory bowel disease (Liu et al. 2015). Association plots for *ERAP1* and *IL23R* are shown the Appendix (Chapter 7, Figure 7.3).

Enrichment with eQTL data

Successfully fine-mapped SNPs listed in Table 5.1 were overlapped with eQTL SNPs from a number of datasets contained within the package XGR (Fang et al. 2016). All significant common eQTL SNPs reported by the authors were retrieved. The background was specified as all GWAS study SNPs, and the enrichment analysis was based on a binomial test for estimating the significance of overlaps using the function *xGRviaGenomicAnno* (Figure 5.2). The top enrichment was with GTEx muscle skeletal, followed by a number of probable unrelated tissues such as GTEx stomach and GTEx aorta. $CD4^+$ T cells were the top enriched immune cell type, followed by GTEx whole blood, $CD8^+$ and NK T cells. A number of monocyte datasets, both stimulated and unstimulated were found enriched, but at a much lower significance. This analysis has a number of limitations, which includes the bias created by missing GWAS loci that were unable to be fine-mapped with the UK Immunochip cohort. Overlap of eQTL and prioritised GWAS SNPs is not an ideal method to identify eQTLs that may be explaining GWAS signals, as these results may be the result of coincidental overlap due to local LD. A more appropriate method is colocalisation analysis, which is able to link eQTLs to GWAS SNPs. Candidate causal eQTLs have been identified using this method in SLE (Odhams et al. 2017), and can be used to prioritise candidate genes and causal variants along with their mechanisms of action.

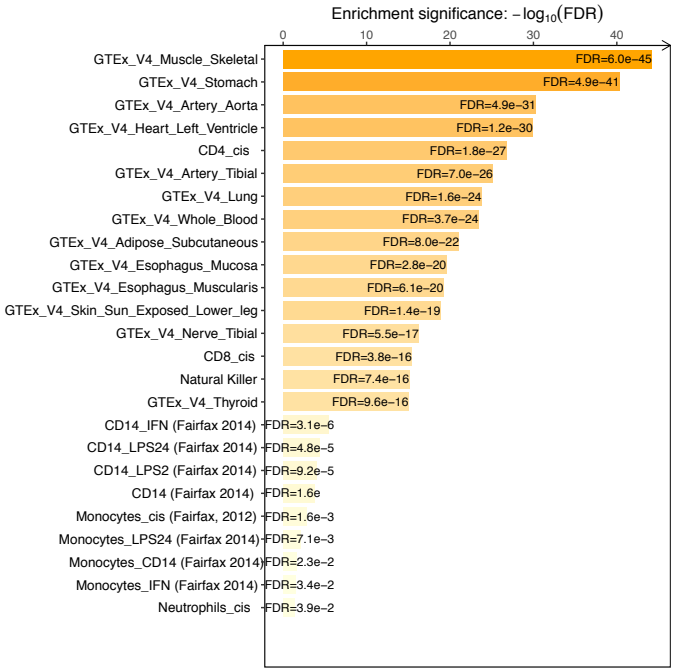


Figure 5.2: Enrichment of fine-mapped SNPs in eQTL datasets. Based on a binomial test for estimating the significance of overlap between common significant eQTLs from GTEx and in house data and all successfully fine-mapped SNPs.

5.3.4 Functional follow up of the *CARD9* locus

Early GWAS and the Immunochip both identified an association with a locus containing the caspase recruitment domain family, member 9 (*CARD9*). Rivas et al carried out a search for low frequency variants at this locus by resequencing, and a rare variant was identified that affected a splice site at exon 11 (Rivas et al. 2011). This splice site variant was found to be protective for AS, CD and UC, in comparison to the common variants identified in various GWAS studies for these diseases (McGovern et al. 2010; Franke et al. 2010; Evans et al. 2011). A variant in LD with the lead associated SNP in AS is a nonsynonymous variant (rs4077515), and creates an amino acid substitution of S12N. The location of this SNP is demonstrated in Figure 5.5. This SNP has been associated with higher *CARD9* expression in Crohn's disease, but it is still unknown whether this coding variant is implicated in the disease pathogeneisis (Franke et al. 2010). SNPs in LD stretch across a number of genes at this locus, including *SNAPC4*, *SDCCAG3*

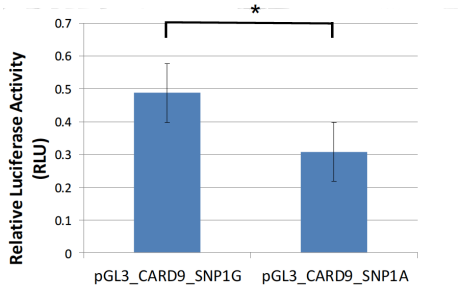
and *SEC16A*, therefore understanding the potentially causal SNPs and the genes implicated in disease at this locus is a current challenge.

This is an example of a functional experiment with the aim to identify the causal mechanism of prioritised SNPs. These experiments were performed within the first year of my DPhil, before fine-mapping for the AS Immunochip loci was implemented. This prioritisation was mainly based on an interest in the *CARD9* gene, and SNPs were chosen based on overlap with various annotation data (Bernstein et al. 2010; Dunham 2012; Adams et al. 2012), along with in house $CD14^+$ eQTL data. An eQTL in $CD14^+$ monocytes for *CARD9* was found in high LD with the SNPs identified in the AS, Crohn's and Ulcerative Colitis GWAS, suggesting that these SNPs are involved in regulating *CARD9* expression in a cell type dependent manner (Fairfax et al. 2014). The peak eQTL SNP in resting and stimulated monocytes for *CARD9* is reported as rs4078099, and SNPs in high LD were also found to be eQTLs for *CARD9* in peripheral blood and NK cells. Overlap with ATAC-seq data suggests that *CARD9* expression is cell type specific when comparing $CD14^+$ monocytes to $CD4^+$ and $CD8^+$ T cells, as shown in Figure 5.5, exhibiting open chromatin peaks in $CD14^+$ monocytes over the promoter.

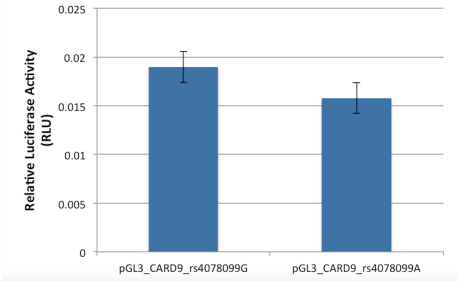
Reporter gene assay

A selection of SNPs prioritised in this region in regards to epigenetic overlap were taken forward to test for enhancer activity using the pGL3 luciferase reporter gene assay, as described in Chapter 2, Section 2.3.3. Allelic constructs were assessed for luciferase activity following transient transfection into HEK293T cells using a lipofection approach. HEK293T cells are human embryonic kidney cells which express the large T antigen which can bind to SV40 promoters. This means they can increase expression in SV40 enhancer vectors, and are stable and reliable to culture and transfect. The region surrounding the SNP was

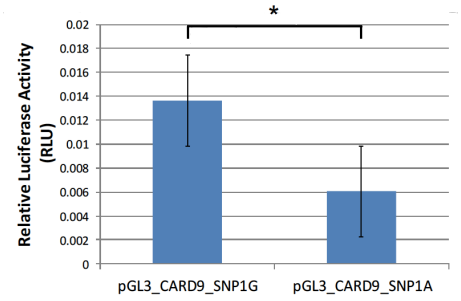
cloned upstream of the luciferase gene, replacing the SV40 constitutive promoter. Rs4078099 is located in the *CARD9* promoter chromatin segmentation marks for K562 cells and primary $CD14^+$ monocytes, and is also overlapping a large number of ENCODE transcription factor binding sites and open chromatin in $CD14^+$ monocytes as shown in Figure 5.5. This construct produced a significant difference in relative luciferase activity between the two alleles in HEK293T cells. In K562 cells, which are an immortalised myelogenous leukemia cell line, no difference was seen between the alleles for rs4078099 in unstimulated cells (Figure 5.3b), while a significant difference was observed on IFN-γ activation. The wild type or G allele again produced significantly higher amounts of luciferase in comparison to the A allele variant, as shown in Figure 5.3(c). K562 cells are relatively similar to monocytes, but may need to be stimulated to be better representative of what may be occurring *in vivo*. This SNP may therefore be associated with decreased *CARD9* expression.



(a) HEK293T cells



(b) K562 cells unstimulated



(c) K562 cells IFN-γ stimulated

Figure 5.3: Luciferase reporter gene activity. Relative light units of luciferase acvitity for 3 replicates between the G and A allele of rs4078099 in a) HEK293T cells b) K562 cells stimulated with IFN-γ for 24 hours. P <0.05 paired t-test (two tailed). Error bars represent standard error.

Allele specific binding in primary monocytes

Using a variety of transcription factor binding data from primary cells, databases and *in silico* predictions (Dunham 2012; Messeguer et al. 2002), candidates for transcription factor binding in both alleles were identified (Table 5.2). ENCODE ChIP-seq Factorbook binding motifs in K562 cells were only identified for the G allele (Wang et al. 2013a), shown in Figure 5.4(a), and there was evidence of a YY1 binding motif to the A allele only (Yant et al. 1995; Messeguer et al. 2002).

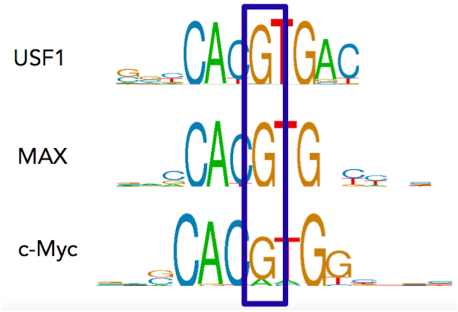
Due to a large number of transcription factor binding predictions for rs4078099, radioactive electrophoretic mobility shift assays (EMSAs) were carried out to further investigate allele specific differences. Radiolabelled probes spanning the binding motifs for this SNP and corresponding to each allele were incubated with nuclear extract from THP1 cells or primary monocytes. THP1 cells were chosen instead of HEK293T or K652 cells due to their increased similarity to primary monocytes. There were a number of DNA:protein complexes formed, as demonstrated by the arrows in Figure 5.4(b). Competition EMSA experiments using molar excess of unlabelled probe corresponding to either allele or an unrelated probe revealed that most of these were likely to be non-specific. Two complexes (II and III) were not effectively competed by the non-specific competitor DNA. It remains unclear from this experiment whether there is evidence of differences in affinity based on competition with each unlabelled allele.

A supershift assay was also performed to elucidate specific transcription factor binding with three transcription factors, YY1, USF1 and c-myc (Figure 5.4c). Antibodies to these proteins were incubated with nuclear extract before the addition of the radiolabelled probes. Binding of the antibody will cause a shift of one band to a higher position in the gel since this complex will migrate more slowly or alternatively abolish protein-DNA binding. A supershift seems to have occurred in compex III for USF1, where the band seems to have shifted to a higher location in the gel. YY1 binding may have abolished complex IV, even though this complex was initially believed to be a non-specific band in the competition EMSA. Both of these events seem to have occurred in both alleles, although ENCODE ChIP-seq data predicts that USF1 will only bind the G allele and the A allele variant creates a binding site fot YY1, therefore this assay was not very informative in respect to allele specific binding.

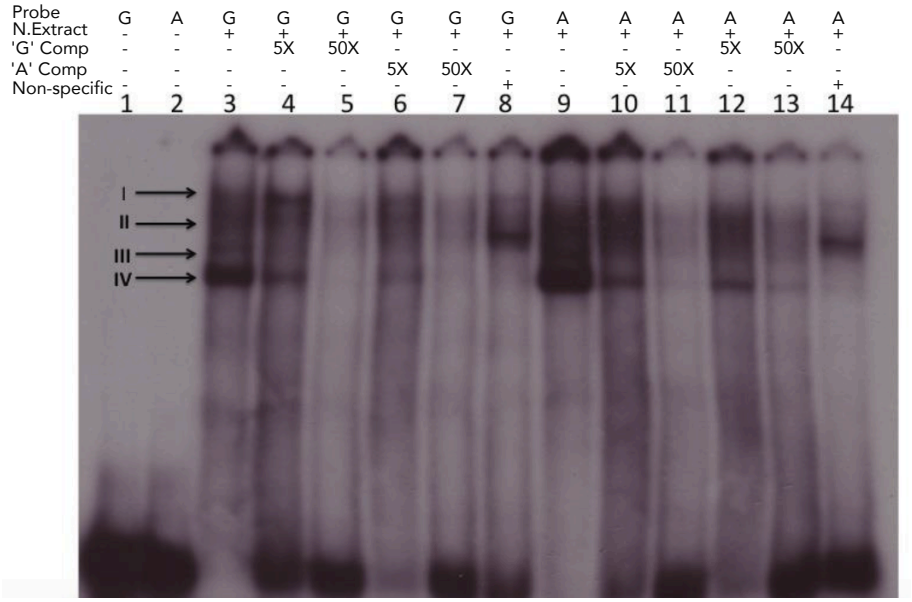
The EMSA using nuclear extract from primary monocytes (Figure 5.4d) reveals different complexes in comparison to the EMSA using THP1 cells and may represent a more accurate picture of what occurs *in vivo*, with one complex in the G allele and two complexes being formed in the A allele. The additional complex in the A allele seems to be competed out with 50X unlabelled probe of both alleles, and is not competed by the non-specific probe, demonstrating that there seems to be an additional binding in the A allele. EMSAs are an in vitro method to detect transcription factor binding, so may not reflect what is occurring *in vivo*. However, these results are encouraging and follow up with an *in vivo* method such as ChIP-seq may yield more consistent and physiologically relevant results.

Sequence	TF	Dissimilarity	String
GAGCAGCCAC**G**TGGTCCTCAG	c-Myc	0.00	CACGTG
	USF1	0.477028	CACGTGGTCC
GAGCAGCCAC**A**TGGTCCTCAG	YY1	0.00	ATGG

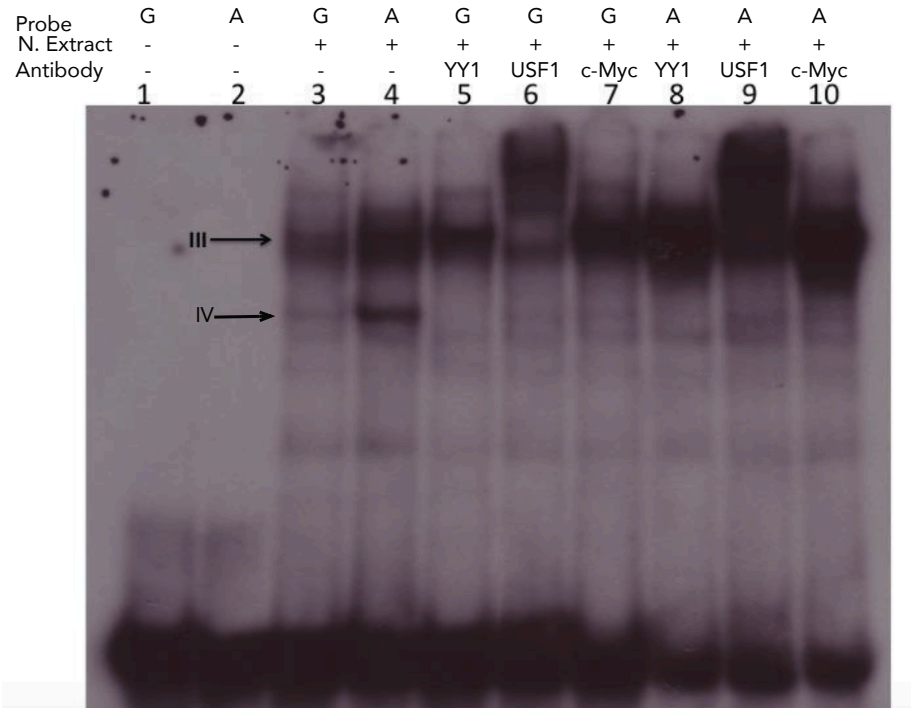
Table 5.2: Transcription factor prediction of the DNA sequence surrounding rs4078099 using PROMO database of putative transcription factor binding sites. c-Myc and USF1 were consistent with the ENCODE ChIP-seq data. The dissimilarity score is a measure of how different the sequence and the string of the binding site are.



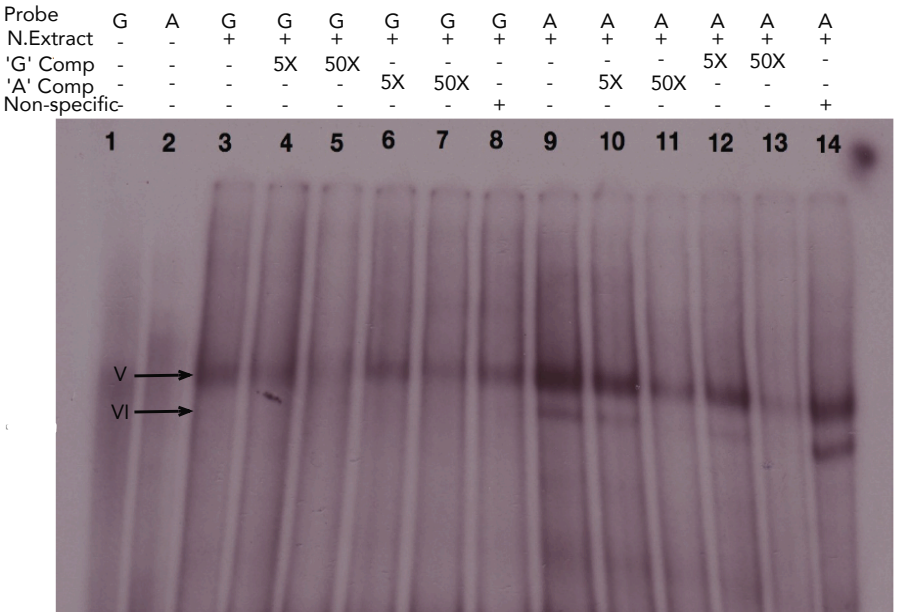
(a) Transcription factor binding motifs, found in K562



(b) Competition EMSA in THP1 cells



(c) Supershift EMSA in THP1 cells



(d) Competition EMSA in primary monocytes

Figure 5.4: Allele specific transcription factor binding. **(a)** Factorbook motifs from ENCODE K652 ChIP-seq data, demonstrating transcription factors binding to the G allele over the A allele at rs4078099. **(b)** Competition EMSA for rs4078099 in THP1 cells. Radiolabelled probes corresponding to the G or A allele of rs4078099 run alone (lanes 1,2) or incubated with nuclear extract from THP1 cells (lanes 3-14) alone (lanes 3,9) or in competition with unlabelled probe (5-fold or 50-fold molar excess of G allele in lanes 4,12 and 5,13 respectively; 5-fold and 50-fold excess of A allele in lanes 6,10 and 7,11 respectively; or an 50-fold excess of an unrelated probe in lanes 8 and 14. **(c)** Supershift EMSA for rs4078099, G or A run with probe alone (lanes 1,2) or incubated with nuclear extract from THP1 cells (lanes 3-10) alone (3,4) or with antibody YY1 (5,8), USF1 (6,9), c-myc (7,10) for the G and A allele respectively **(d)** Competition EMSA with rs4078099 with probes alone (1,2), incubated with primary monocytes nuclear extract (3,4) or in competition with unlabelled probe (5 fold or 50 fold molar excess of G in 4,12 and 5,13 respectively; 5 fold or 50 fold molar excess of A in 6,10 and 7,11 respectively). 50 fold excess of unrelated probe in 8,14.

Fine-mapping analysis at the *CARD9* locus

Fine-mapping was later carried out at the *CARD9* locus, and found multiple independent signals demonstrating high LD, and the lead SNP from the Immunochip study rs1128905 alone in one of those signals. After further interrogation with the original IGAS cohort it was found that the association at rs1128905 was a genotyping error (Appendix, Chapter 7, Figure 7.4). This may have occurred due to cases and controls being genotyped separately in different research centres, causing some differences in SNP calling. After removing the

SNP and re-imputing it, the primary signal was found to be tagging the original SNPs identified through earlier GWAS (Evans et al. 2011), and in high LD with the SNPs investigated in this functional follow up work. My candidate SNP rs4078099 was found in the European cohort credible set, with a low posterior probability of 0.01. The OR at the *CARD9* locus is relatively small (1.1), meaning that the fine-mapping resolution is not very precise. The fine mapped SNPs from the European cohort are demonstrated in Figure 5.5, with my candidate SNP rs4078099 labelled in green. There are 27 SNPs in the 95% credible set, with the majority appearing to be non functional when overlapped with annotation data. rs4078099 overlaps ATAC-seq peaks in $CD14^+$ monocytes only, therefore may be important due to its high cell specificity and its location in the gene promoter. A variant overlapping a T cell specific peak in *SDCCAG3* (rs1051957) falls into an active enhancer site in GM12878, and primary $CD4^+$ and $CD8^+$ T cells. It also overlaps with $CD4^+$ and $CD8^+$ digital genomic footprinting sites, demonstrating functional activity and the binding of TFs. One fine mapped SNP, rs4307445 is overlapping a potentially differentially open chromatin site in $CD8^+$ T cells, with a peak found in 3 healthy controls and no AS patients (data not shown), and is also overlapping transcription factor binding sites from ENCODE and ChromHMM enhancer sites in $CD4^+$ and $CD8^+$ T cells. This SNP interacts with the gene *INPP5E* in total activated $CD4^+$ T cells in promoter capture Hi-C data (CHiCAGO score = 5.56) (Javierre et al. 2016; Schofield et al. 2016), therefore SNPs in this association signal may be acting through multiple genes and mechanisms at this locus.

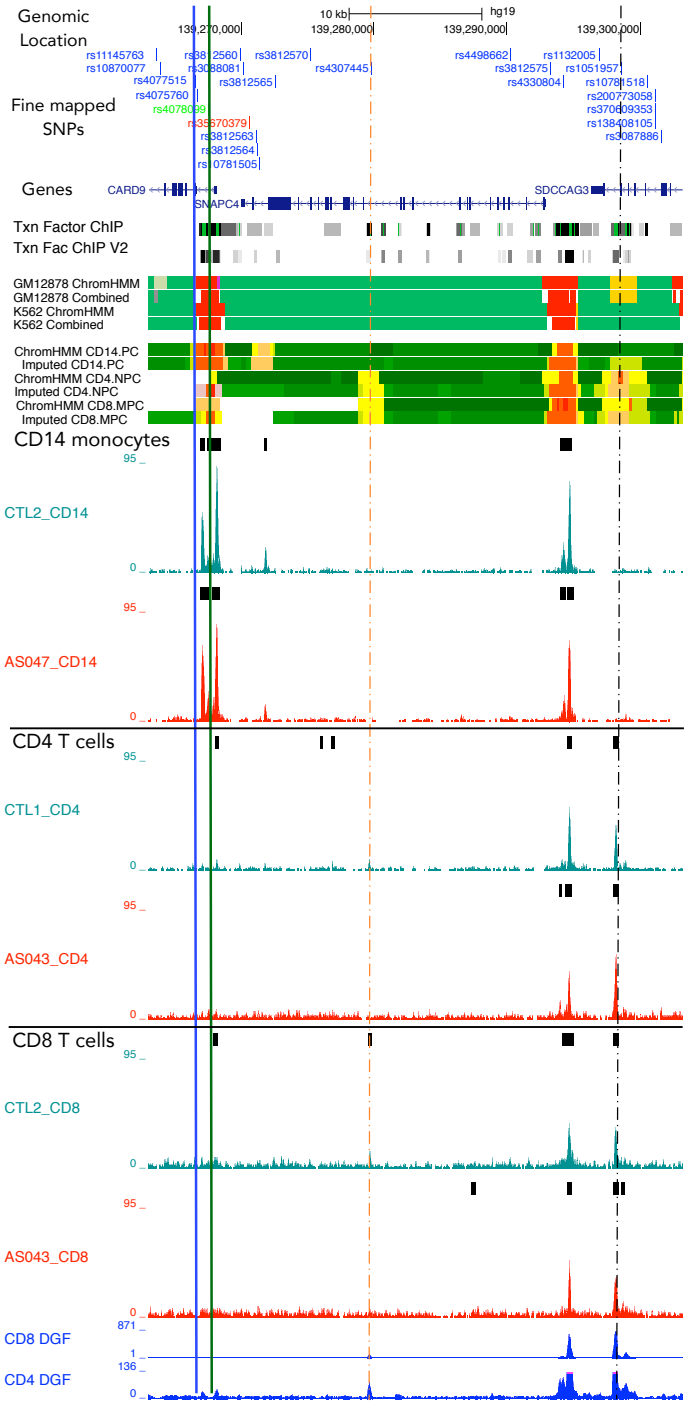


Figure 5.5: Functional annotation of fine mapped SNPs at the *CARD9* locus. UCSC Genome Browser view of fine-mapped SNPs, along with ATAC-seq read density distribution showing an example of a healthy control sample (green) and an AS patient (red), in $CD14^+$ monocytes, $CD4^+$ and $CD8^+$ T cells. The SNP with the highest posterior probability is highlighted in red, but does not overlap open chromatin in any cell type. The nonsynonymous variant rs4077515 is highlight in blue, and rs4078099 is shown in green, demonstrating overlap with open chromatin in $CD14^+$ only plus many transcription factors. Fine-mapped SNPs stretch across 4 genes (2 SNPs in *SEC16A* are not shown in this figure), and the SNP rs1051957 in *SDCCAG3* falls within an active enhancer in GM12878 cells, and overlaps open chromatin for $CD4^+$ and $CD8^+$ T cells only, as demonstrated by the black dashed line. The orange dashed line demonstrates a potential differentially open region in $CD8^+$ T cells with rs4307445, where a peak is found in 3 healthy control samples, and in no AS patients.

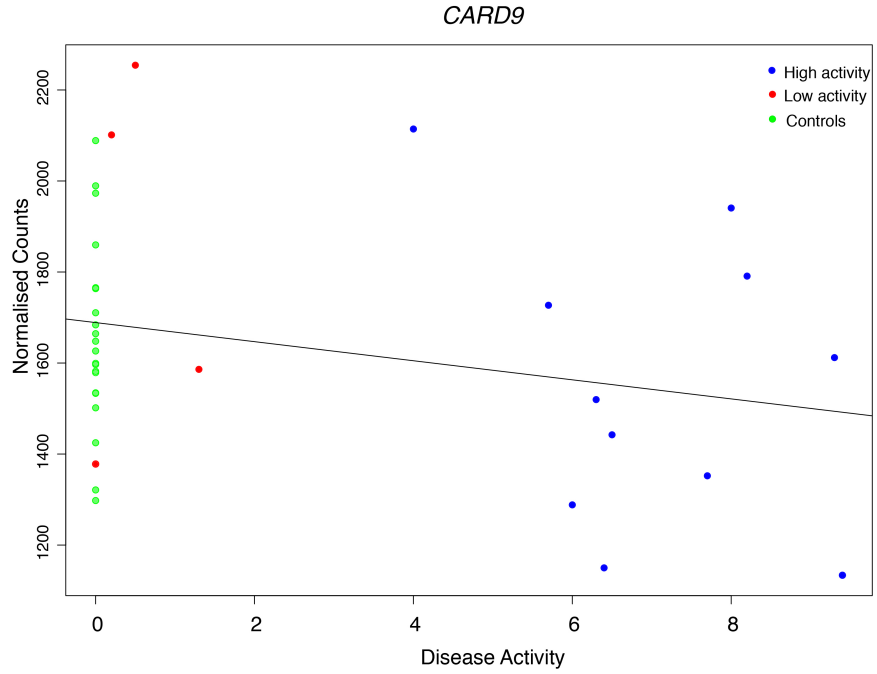
Gene expression data for *CARD9* in AS

RNA-seq gene expression results in $CD14^+$ monocytes between healthy controls and AS patients (Chapter 3) did not find a significant difference in the expression of *CARD9* (Log2FoldChange=0.174, FDR=0.096). The smallest adjusted *P*-value was found in $CD14^+$ monocytes in comparison to the other 5 cell types, where the FDR values were extremely large. This could suggest that *CARD9* may have the potential to reach significance in a higher powered study, or by including AS patients that also exhibit some overlap with gut inflammatory diseases, since this association is shared with both CD and UC. The genes involved in the NOD-like receptor signalling pathway, believed to have a potential role in the defence against extracellular bacterial, are shown in the manually curated KEGG pathway, demonstrating genes up or down-regulated for AS patients (Appendix, Chapter 7, Figure 7.1). The risk allele in AS is the A allele, therefore we expect AS patients to have decreased *CARD9* expression instead of increased as these results demonstrate.

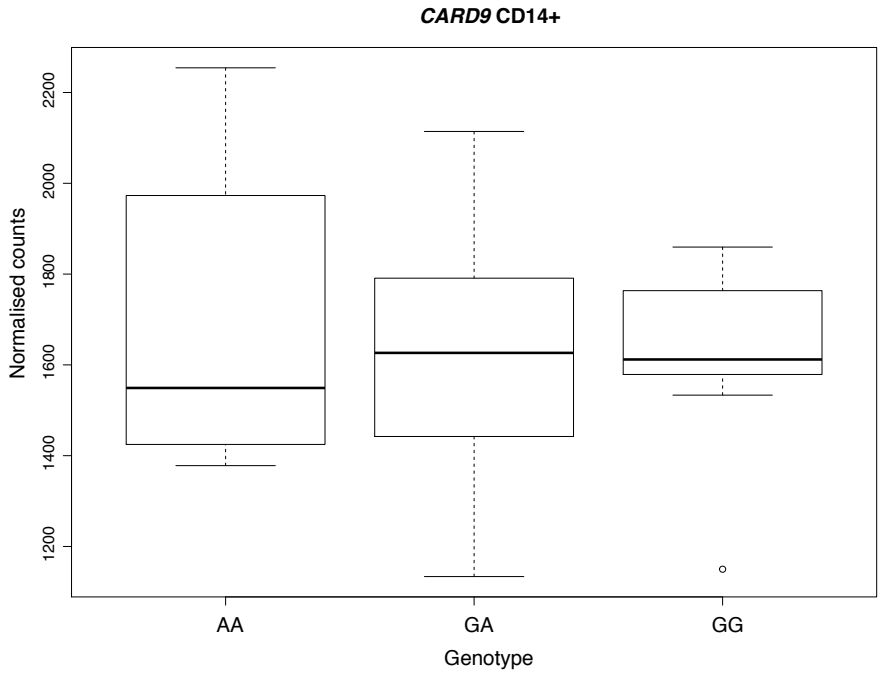
Plotting the normalised counts against disease activity does, however, show a very weak trend; AS patients with the highest disease activity exhibit the lowest *CARD9* normalised counts, which may demonstrate a potential link with *CARD9* expression and disease activity (Figure 5.6). Using the RNA-seq data to obtain the genotypes for each individual and plotting these against the normalised counts demonstrated no significant difference in genotype at this SNP. Rs4078099 may not be the functional SNP regulating *CARD9* expression, and it may also be possible that *CARD9* may also not be the AS associated gene at this locus since it did not reach significant in any of the cell types investigated with RNA-seq. The allele specific expression of *CARD9* may have more relevance in tissues of the intestinal mucosa, and therefore further investigation is required to understand whether *CARD9* and the NOD-like signaling pathway has a relevance in AS pathogenesis. *CARD9* was found significantly up-regulated in treated

AS patients with anti-TNF-α therapy compared with untreated AS patients (FDR=0.02) (Chapter 4). These patients were treatment responders, therefore reduced disease activity again correlated with increased *CARD9* expression.

INPP5E was only found to be close to significance in CD4+ T cells (Log2FoldChange=0.315, FDR=0.078), suggesting this gene may be of importance specifically in this cell type, as also suggested by the promoter capture Hi-C chromatin interaction data at rs4307445. There was no differential expression of *SDCCAG3*. These results may indicate that various SNPs and genes may be relevant in different cell types, demonstrating the complexity at GWAS loci.



(a)



(b)

Figure 5.6: *CARD9* normalised counts with **(a)** Disease activity based on BASDAI score. The lowest counts are found in high activity AS patients, therefore *CARD9* may relate to disease severity. **(b)** rs4078099 genotype specific expression. No effect of genotype is seen on *CARD9* expression.

5.3.5 GWAS overlap with differential open regions from ATAC-seq

We are unable to detect differential open chromatin regions (DOCs) which overlapped with the fine-mapped GWAS SNPs by the method demonstrated in Chapter 4. This may be due to the small sample size in the AS ATAC-seq cohort. However, by interrogating the fine mapped Immunochip loci with the ATAC-seq data through the UCSC genome browser, I was able to further fine-map SNPs with functional relevance in a cell type dependent manner. Functional annotation is important in further refining the functional mechanism that may be acting at that locus, as statistical methods may still produce a large number of prioritised SNPs. Overlap will also give an indication of the cell type most relevant in that locus, which may have potential implications in understanding the general aetiology of AS. An example of cell type specific fine-mapped SNPs rs887477 and rs11064154 at the *TNFRSF1A-LTBR* locus, overlaps an ATAC-seq peak in $CD14^{+}$ monocytes only, as demonstrated in Figure 5.7. I will now present an example of integrating ATAC-seq data, fine-mapped SNPs, RNA-seq data and publically available data to generate a hypothesis for an AS GWAS locus in a differentially closed chromatin region.

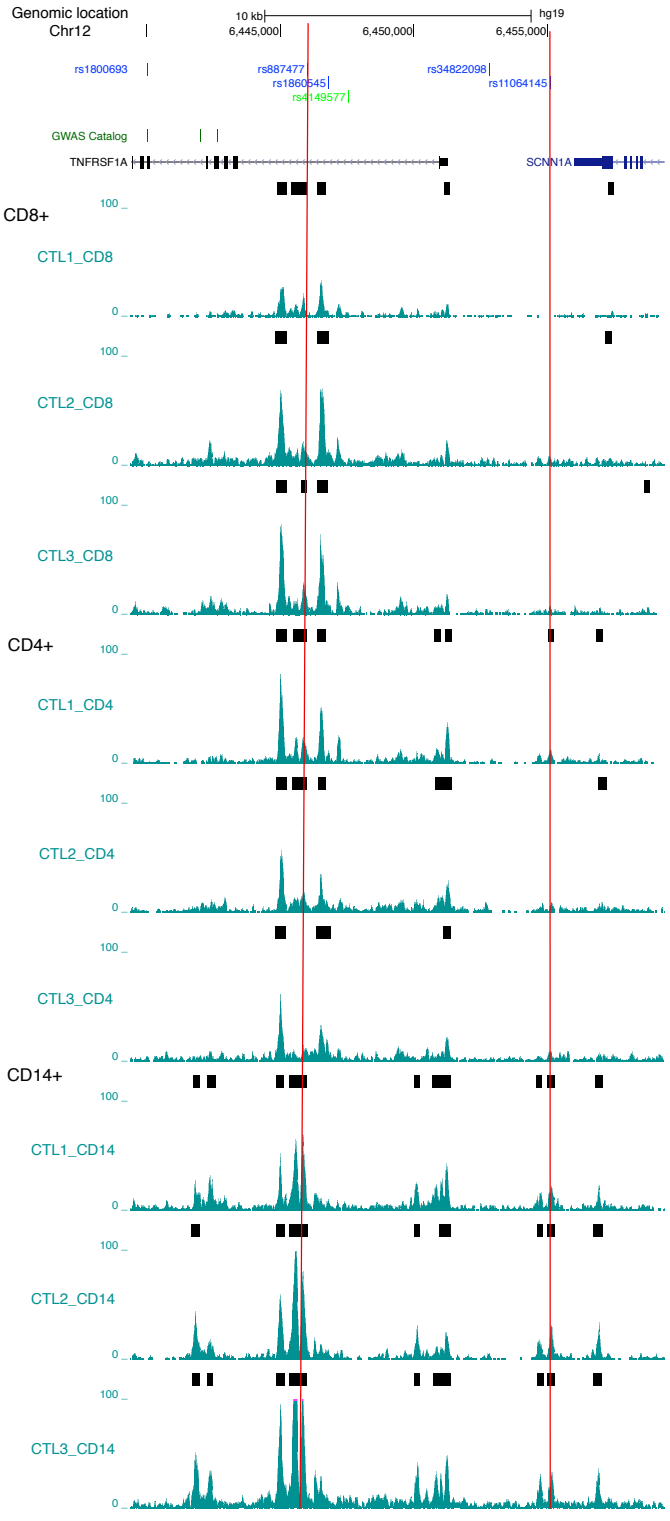
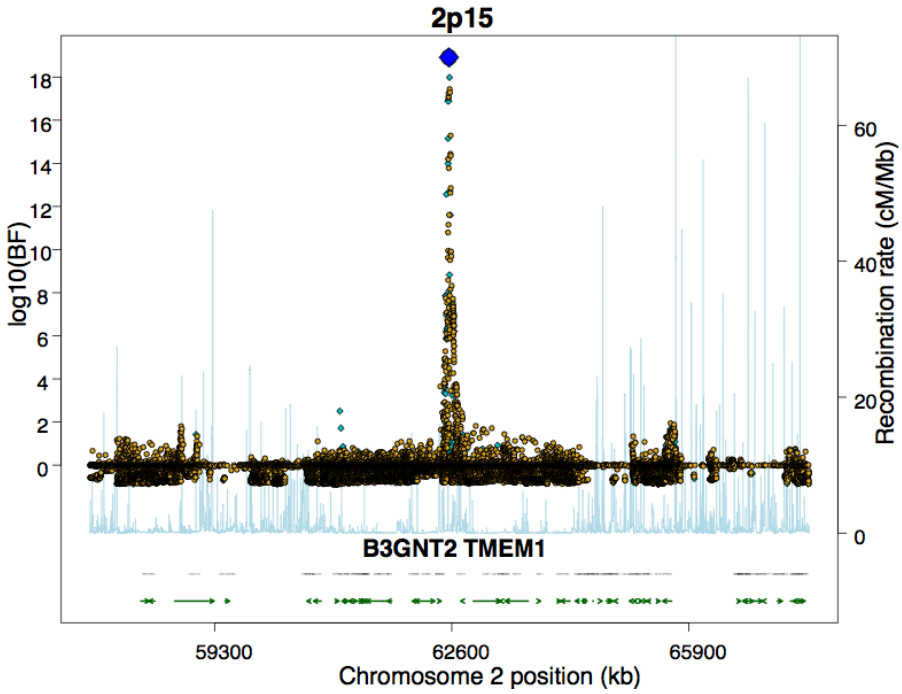


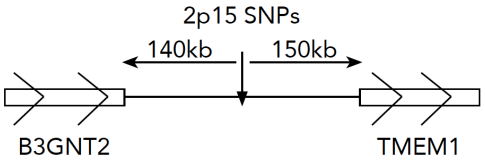
Figure 5.7: Fine-mapped SNP at the *TNFRSF1A-LTBR* locus, demonstrating open chromatin at rs887477 and rs11064145 in CD14$^+$ monocytes ATAC-seq data. UCSC Genome Browser view of fine-mapped SNPs, along with ATAC-seq read density distribution showing examples of a healthy control samples in the 3 cell types. The SNPs highlighted in red, especially rs11064145 may be functional in CD14$^+$ monocytes only.

Locus 2p15

The association with AS at locus 2p15 is found in a gene desert, therefore may be acting as a long-range enhancer to regulate gene expression. There is no LD with any coding variants at this locus, therefore the association is not tagging any non-synonymous mutations in nearby exons. As previously mentioned, refinement at this locus was extremely successful, with 3 SNPs explaining 95% of the signal (Figure 5.8a). Rs4672505 explained roughly 60% of the posterior probability, with the remaining two SNPs explaining 40% and therefore is likely to be the causal SNP at this locus. This signal is also a shared susceptibility locus with psoriasis (Tsoi et al. 2012).



(a)



(b)

Figure 5.8: Refinement of the 2p15 locus to 3 SNPs in the credible set. **(a)** Fine-mapping at locus 2p15. The association at this locus has a log10 bayes factor of 18.42. The SNPs highlighted in blue form part of the credible set and are assigned posterior probabilities. **(b)** The location of the locus 2p15 SNPs in relation to the closest genes. The associated SNPs lie in a gene desert, between genes *B3GNT2* and *TMEM1*.The blue lines represent recombination hotspots.

When viewing this locus with our ATAC-seq data, the lead SNP rs4672505 overlaps a peak in $CD8^+$ T cells only, suggesting that chromatin accessibility at this SNP is cell type specific (Figure 5.9). The AS patients all presented a peak overlapping this SNP in $CD8^+$ T cells, except for one patient GNOPBAS043, in which the peak was absent (Figure 5.10). The peak was again absent in the post 3 month TNF-α inhibitor therapy sample, providing some evidence that this was not an artefact of the experiment. The peak was also absent at the same

location in a psoriasis $CD8^+$ patient sample collected by a colleague. The SNPs in both diseases are in high LD (Tsoi et al. 2012), therefore we could hypothesise a shared causal mechanism between the two diseases at this locus. Many AS patients often exhibit symptoms of psoriasis, or have close family members that do, also confirming a link between the two diseases. All healthy control samples demonstrated a peak at this location.

Figure 5.9: ATAC-seq chromatin accessibility peaks showing cell type specificity at rs4672505 for CD8$^+$ T cells. UCSC Genome-browser view of control samples showing the ATAC-seq read density distribution surrounding the SNP. CD8$^+$ T cells are the only cell type with open chromatin at the SNP. The small peaks in some CD4$^+$ T cells may be coming from the Th17 subset in these cells, as Th17 also demonstrate open chromatin at this SNP in ENCODE data.

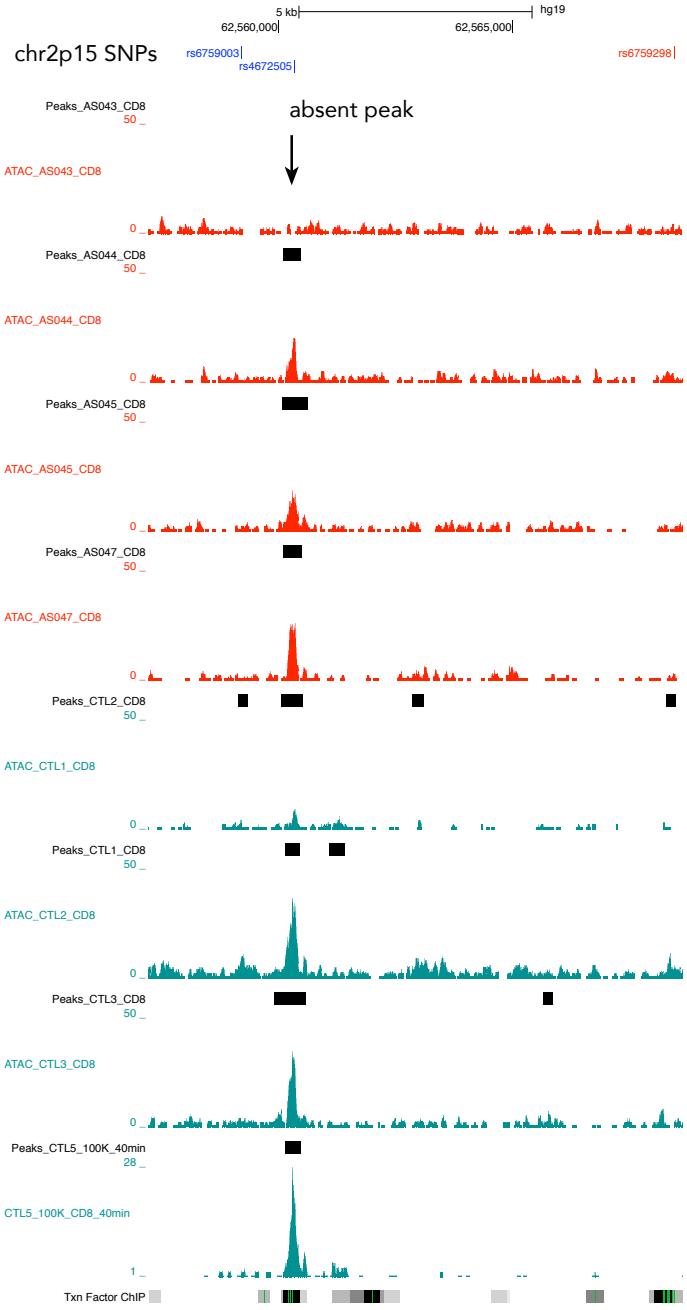


Figure 5.10: GNOPBAS043 exhibiting closed chromatin accessibility in CD8$^+$ T cells at rs4672505. UCSC Genome-browser view of the ATAC-seq read density distribution surrounding the SNP in AS patients (red) and healthy controls (green). All samples have open chromatin, apart from one patient sample GNOPBAS043. ENCODE transcription factor ChIP-seq is also shown demonstrating a variety of transcription factors binding around rs4672505.

Allelic imbalance in ATAC-seq reads at rs4672505

To determine the genotype for all patients and controls, sanger sequencing was performed in the region surrounding the SNP. The patient GNOPBAS043,

demonstrating closed chromatin, was found to be homozygous for the A allele, and all other samples were either heterozygous or homozygous for the G allele. Chromatin accessibility may therefore be inactive in the A allele, as patient GNOPBAS043 has zero reads mapping at this location in ATAC-seq, since the Tn5 transposase is unable in access the DNA in this genotype. To test this hypothesis, we can check for allelic imbalance in the ATAC-seq reads in the heterozygous individuals. All 4 heterozygous individuals had 100% of their ATAC-seq reads mapping to the G allele only (Figure 5.11), suggesting that chromatin is indeed closed in individuals possessing the A allele at rs4672505. This is a genotype specific effect rather than a patient-control effect in our data, as our current sample size does not enable us to explore whether there is an AS specific effect at this SNP. From the GWAS and the fine-mapping analysis we can assume that this may be the causal mechanism driving the association at this locus.

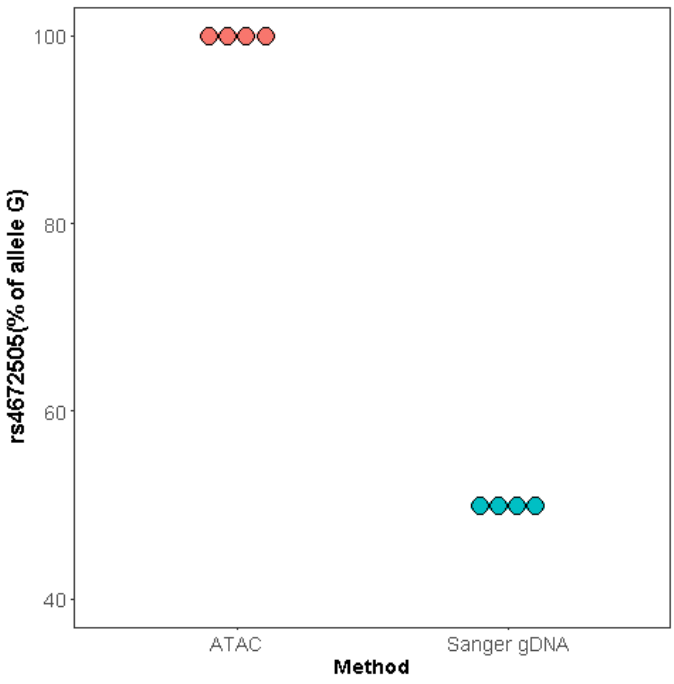


Figure 5.11: Allelic imbalance of ATAC-seq reads at rs4672505 in heterozygous individuals. In a heterozygous individual you would expect 50% of the reads to come from G and 50% from A. All ATAC-seq reads in these heterozygous individuals are from the G allele only.

Chromatin conformation capture with Promoter Hi-C data at rs4672505

To identify a possible mechanism of action, plotting promoter capture Hi-C data from rs4672505 revealed a high confidence interaction with the promoter of a gene downstream called *B3GNT2* (Figure 5.12). Rs4672505 falls into a strong enhancer chromatin segmentation mark in GM12878, but a repressive chromatin segmentation mark in primary $CD8^+$ T cells from Roadmap data (Bernstein et al. 2010) (Figure 5.12). It could be the case that these primary $CD8^+$ T cells from Roadmap are also heterozygous for the A allele, and therefore are exhibiting repressive or closed chromatin. ChIP-seq experiments from ENCODE in GM12878 suggest a whole variety of transcription factors binding around the SNP, of which the strongest was RUNX3, which is also an AS associated gene. Transcription factor binding predictions using PROMO also suggest binding motifs over the SNP for transcription factors such as c-Ets-1, GR-α, STAT-β, IRF-1 and HNF-3α (Messeguer et al. 2002). Which transcription factor is binding in $CD8^+$ T cells at this position requires further investigation.

Further evidence for the chromatin interaction with *B3GNT2* is provided by the GTEx eQTL study. Although in whole blood, the study reports an eQTL for SNPs in LD with rs4672505 (FDR= $7.3\text{x}10^{-31}$) for *B3GNT2*. A methylation-QTL (meQTL) has also been reported in a CpG island cg04342863 200kb away from the SNP in individuals with the A allele only (FDR= $3.27\text{x}10^{-31}$), suggesting that the variant may be responsible in repressing the chromatin landscape (Bonder et al. 2017).

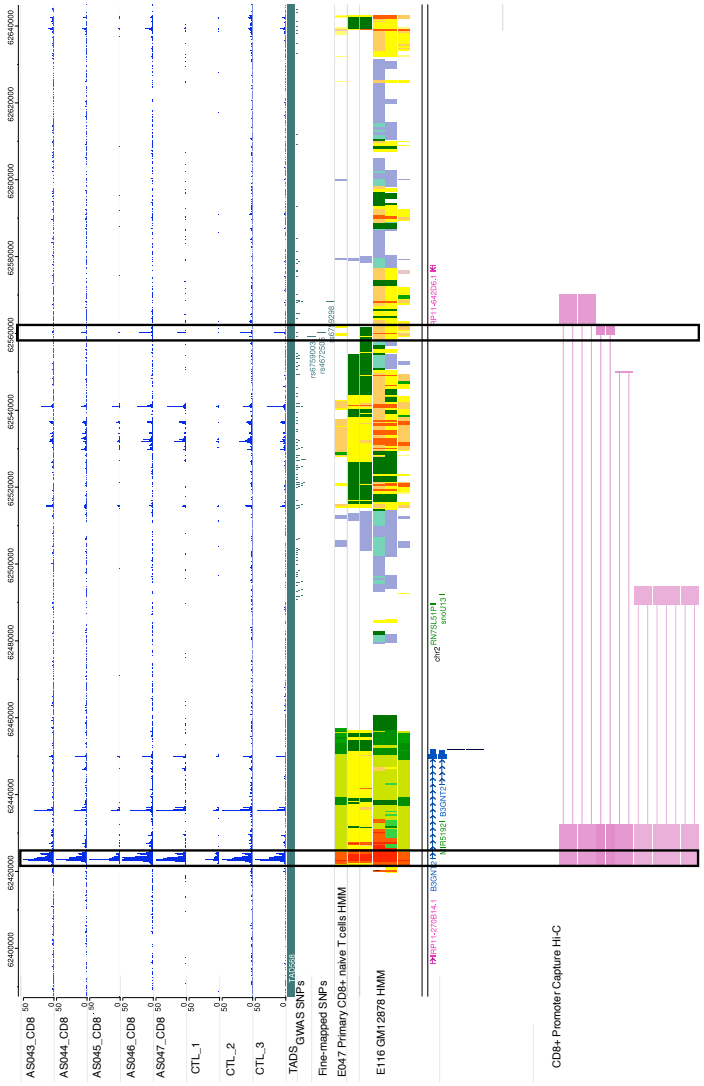


Figure 5.12: rs4672505 interacts with the promoter of *B3GNT2* in CD8$^+$ T cells. This is a high confidence interaction from the promoter capture Hi-C data for CD8$^+$ T cells between rs4672505 and rs6759008, two SNPs in the credible set and found within the same restriction fragment with *B3GNT2*. These SNPs fall into an enhancer chromatin segmentation mark in GM12878 as displayed by the orange segment, however in primary CD8$^+$ T cells it is a repressive mark, shown here in green.

B3GNT2 gene expression

B3GNT2 was found differentially expressed in the RNA-seq results from Chapter 3 in $CD14^+$ monocytes only (FDR = 0.048), and is down-regulated in AS patients but with a very low log2 fold change of -0.26. To investigate whether closed chromatin, or the A allele at rs4672505 is related to *B3GNT2* expression in $CD8^+$ T cells, the RNA-seq normalised counts were plotted against genotype, as shown in Figure 5.13. Genotype is correlated with *B3GNT2* expression in $CD8^+$ T cells only, with the A allele, or closed chromatin demonstrating the lowest expression and the G homozygotes exhibiting significantly higher gene expression in $CD8^+$ T cells in comparison with the other two cell types. This may suggest that rs4672505 is involved in regulating *B3GNT2* in $CD8^+$ T cells only, and that other variants may be involved in regulating expression in $CD14^+$ monocytes and $CD4^+$ T cells. The cohort only contains one individual that is homozygous for the A allele, therefore more individuals with this genotype are required to investigate this association.

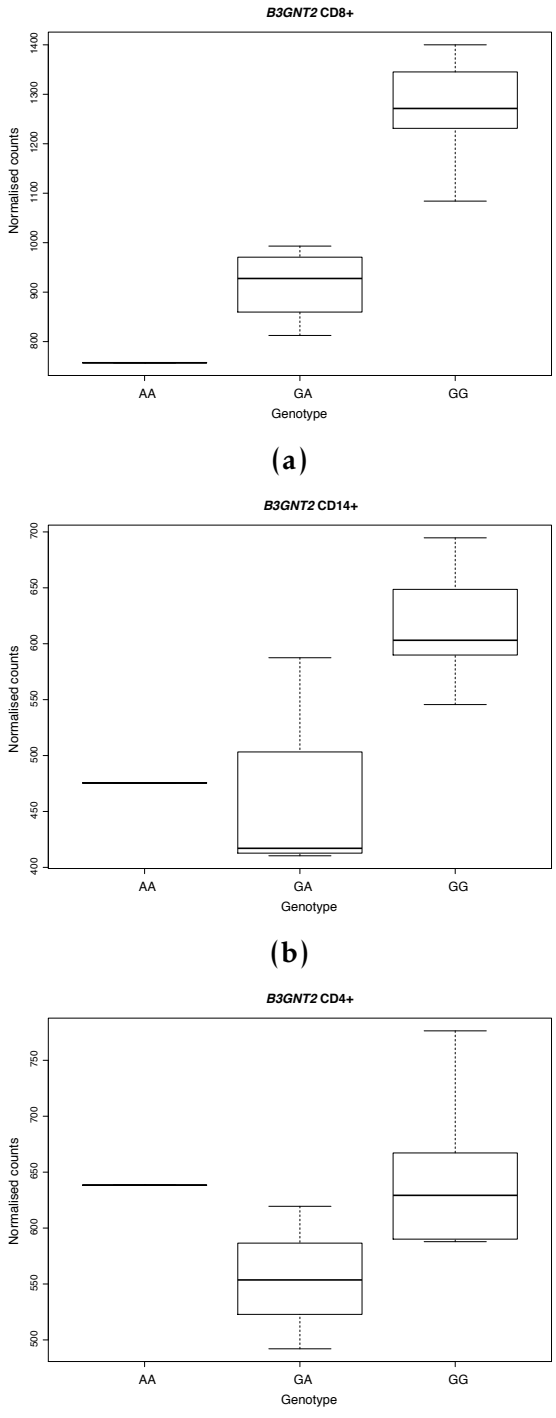


Figure 5.13: Boxplots of genotype specific *B3GNT2* expression in CD8$^+$ T cells only **(a)** *B3GNT2* normalised counts in CD8$^+$ T cells demonstrate a trend in expression, with the lowest expression in the A homozygote, the AS patient with closed chromatin at that location. **(b)** *B3GNT2* normalised counts in CD14$^+$ monocytes, and **(c)** CD4$^+$ T cells do not show a correlation with genotype, demonstrating a possible cell type specific effect of the SNP in CD8$^+$ T cells only.

5.4 Discussion

The main results of this chapter included fine-mapping of the AS susceptibility loci from the Immunochip study using the UK cohort, enabling the prioritisation of loci and regulatory variants for future functional studies. This was able to further refine the functional variants of the *CARD9* locus, and the highly prioritised variant rs4672505 at the chromosome 2p15 locus, and can identify patient and cell type specific chromatin accessibility at GWAS loci.

5.4.1 Variant prioritisation

The key challenge of understanding GWAS signals is the prioritisation of causal variants in regulatory DNA elements to determine the functional mechanisms governing its potential effects on disease. This understanding will be critical in the road to identify novel therapeutic approaches by characterising the biological networks responsible for disease pathogenesis and phenotype. The results from this chapter demonstrate that the integration of ATAC-seq data in specific immune cell types provides an effective way of highlighting potential molecular mechanisms at AS GWAS loci.

The identification of candidate causal variants is often complicated by multiple association signals at a single locus, as well as LD structure which could implicate hundreds of SNPs at a single locus. The accurate identification of potentially causal variants are required for integration with cell type specific data and in the design of well informed functional follow up studies to determine pathogenic mechanisms. Our fine mapping approach is a bayesian method using dense and imputed genotyping data, which generates a credible set of SNPs that is likely to contain the causal variant (Maller et al. 2012). This approach assumes that there is just one causal variant at a locus, and

the conditional stepwise approach to identify additional variants assumes independent association signals in the region. The limitations to this method may therefore produce incorrect assumptions and results. The ability to account for multiple causal variant requires the consideration of multiple SNP models, which can be computationally limiting. Methods using bayesian stochastic search approaches can be further investigated to over come these limitations (Wallace et al. 2015). The fine-mapping of AS associated loci using a Bayesian statistical approach for the Immunochip study has enabled us to reduce down the number of candidate SNPs at each locus. Further limitations to our current fine-mapping approach include sample size, as the analysis was carried out with a subset of the Immunochip genotyping data. This reduced power resulted in a large number of loci unable to be effectively fine-mapped, which were successfully fine mapped in the European cohort (Bunt et al. 2015). Comparisons with fine mapping results from the European AS cohort occasionally lead to some variations in prioritised SNPs, reinforcing the fact that results should always be interpreted with all available functional data, since the output can vary depending on the input samples and the specific population. Fine-mapping resolution was also limited on loci with smaller effect sizes, and additional genotype information from imputation does not add resolution to a cohort of this size. For smaller effect sizes, much larger sample numbers are required, but if there is extended and strong LD at a locus, large sample sizes may still be uninformative (WTCCC 2007). The ranking of prioritised SNPs by posterior probabilities may also complicate comparisons on studies ranked by *P* values. In the future, deep resequencing of all disease associated loci may be cost effective and informative for identifying low frequency variants that may not be detected in GWAS.

Approximate bayesian fine-mapping of the 26 AS Immunochip loci was able to reduce 10 of the 26 loci to fewer than 10 SNPs in the 95% credible set. 19 of the 26 credible sets included the lead variant reported in the lead GWAS

analysis (Cortes et al. 2013). The *ERAP1* association is thought to be limited to HLA-B*27 positive individuals via epistasis, which suggests that its pathogenesis (rs30187; Arg528Lys) is mainly related to peptide trimming in the endoplasmic reticulum and antigen presentation by MHC Class I molecules (Evans et al. 2011; Cortes et al. 2013). However, it has also been demonstrated that the effect of a 3 SNP haplotype may be affecting *ERAP1* gene expression in antigen presenting cells (Costantino et al. 2015). mRNA levels are correlated with protein abundance and enzymatic activity, but it is still unclear how changes in EPAP1 enzymatic activity is contributing to AS pathogenesis. Kadi et al, found the nonsynonymous SNP rs30187 is associated only with AS and age of onset and not with non radiographic forms of SpA, therefore it has been suggested that this SNP may be associated with severity of disease (Kadi et al. 2013). The *IL23R* locus lead prioritised variant using the fine mapping approach was the INDEL 1:67699535:GTT:G. Unpublished fine-mapping analysis in the lab also found the same INDEL when fine-mapping the psoriatic arthritis SNPs, suggesting a shared common primary signal between diseases, which is tagging the primary SNP (Cortes et al. 2013). The secondary signal at the *IL23R* may instead be acting through the *IL12RB2* gene.

5.4.2 Locus 2p15

Currently the power to annotate GWAS hits with patient specific ATAC-seq data is low due to the small sample size in our cohort. A differential open peak was identified in one GWAS hit, and had a strong genotype specific effect that resulted in closed chromatin in the variant allele. Only one patient exhibited this effect and therefore larger sample sizes in our ATAC-seq cohort are required to further understand whether this phenomenon is over represented in AS, and whether this is a causal variant in a subset of AS patients. The overlap in GWAS hits between immune mediated diseases is now well established, with family

members of the affected at an increased risk of a number of immune mediated diseases. This is already clearly demonstrated in our AS cohorts in Chapter 3 and Chapter 4, with a majority of patients exhibiting overlap with psoriasis, RA and CD. The 2p15 locus is also a GWAS hit in psoriasis, and differences in genotype specific chromatin accessibility was also found in a psoriasis patient, further enforcing the potential importance of this SNP in these diseases. Whether there is allele specific binding of transcription factors at this SNP is currently being investigated within the group, along with chromatin conformation capture using Capture-C in $CD8^+$ T cells.

The current working model for establishment and maintenance of active enhancers is that transcription factors bind to the DNA, position the nucleosomes, and then serve to keep the region between the nucleosomes in an open conformation. A SNP may be disrupting a transcription factor binding motif for these transcription factors, causing an allele specific effect in chromatin accessibility (Spitz and Furlong 2012). These SNPs may also have an architectural role in disrupting the chromatin conformation contacts between enhancers and promoters. Rs4672505 may be inhibiting the binding of a transcription factor that serves to keep the chromatin in an open conformation.

Linking a regulatory variant to its target gene is a challenge, due to the experimental limitations of using the correct cell type and environmental status, as well as the interpretation of results due to the large numbers of false positives which may occur in chromatin conformation capture methods. Promoter Capture Hi-C data in $CD8^+$ T cells identified a high confidence interaction with *B3GNT2*, which encodes a type II transmembrane protein. It is a member of the beta-1,3-N-acetylglucosaminyltransferase or glycosyltranferase family and is highly expressed in skin, small intestines and colon. It is involved in the synthesis of sialyl-Lewis X, which is essential for binding to selectins, which aid in the migration of T cells into target tissues. Some congenital disorders have

recurring severe infections stemming from an absence of Sialyl-Lewis X attached to E-selectin ligands on their neutrophils. Sialyl-Lewis X is currently being researched for detection and treatment of immune disorders due to its presence on leukocytes and novel anti-inflammatory therapies, such as putative inhibitor of poly-N-acetyllactosamine biosynthesis, which affects selectin ligand activity has demonstrated efficacy in a rodent skin inflammation model (Zollner and Asadullah 2003). Another study, which generated knock out mice for *B3GNT2* found reduced number of polylactosamine structures, which are produced by the polylactosamine synthase B3GNT2. Hyperactivation of lymphocytes occurred due to a lack of polylactosamine on receptor molecules in *B3gnt2-/-* mice, and indicates that polylactosamine has an important role in immunological functions (Togayachi et al. 2010). This may provide some evidence that the dysregulation of *B3GNT2* may have implications in multiple immune mediated diseases and its potential as a therapeutic target. *B3GNT2* gene expression was lower in the A allele in $CD8^+$ T cells only, but whether rs4672505 has an effect on chromatin conformation is yet to be determined.

5.4.3 Functional work at the *CARD9* locus

Functional follow up studies for prioritised SNPs are important in understanding the causal mechanism that may underlie the disease association at a particular locus. The *CARD9* locus has extensive LD, with associated SNPs stretching over 150kb. A cell specific eQTL was demonstrated in naive monocytes for *CARD9*, which overlapped the GWAS signal in AS and Crohn's disease (Fairfax et al. 2014). The most promising SNPs in high LD were investigated, providing a candidate in the *CARD9* promoter which overlapped many transcription factor binding motifs. Allele specific expression was investigated using a luciferase reporter gene assay, which demonstrated a significant difference between the G and the A allele in HEK293 T cells and INF-γ stimulated K562 cells. The

K562 activated state may be more similar to what occurrs in primary monocytes, as a significant difference in expression was not demonstrated in transfections with naive K562 cells. A secondary independent eQTL with IFN-γ stimulation was found in primary monocytes for *CARD9* after conditioning for the original eQTL signal, suggesting that there may be a context specific effect also acting on *CARD9* expression (Fairfax et al. 2014). SNPs in this signal are found 150kb downstream of the *CARD9* promoter, and were found to be protective for Crohn's disease (Fairfax et al. 2014). Although these SNPs are not found in the independent signals from the fine-mapping analysis in AS, there may be a similar inducible mechanism occurring upon IFN-γ stimulation in $CD14^+$ monocytes. The European cohort fine-mapping results also had two additional independent signals, but neither reached genome-wide significance, and contained hundreds of SNPs, suggesting there was not enough power to generate these credible sets.

The increased luciferase expression associated with the G allele may suggest binding of a transcriptional activator such as MYC:MAX or USF1, both shown to bind preferentially to the G allele only, as shown in ChIP-seq experiments (Dunham 2012) or *in silico* predictions using PROMO. The A allele creates a binding motif for YY1 as demonstrated in 5.2, which is a transcriptional repressor. YY1 also directs histone deacetylases to the promoter, and this may explain the lower expression in the A allele due to repressed gene expression. Reporter gene assays have several limitations, including technical challenges in transfection efficiencies and the unavailability of suitable cell lines. Using a plasmid may also not demonstrate what may be occurring in the true cellular environment. CRISPR/cas9 genome editing has revolutionised the potential to study functional variants, as genomic modifications can be introduced within the genome of a cell type of interest. Issues with transfection efficiencies can be overcome by using human pluripotent stem cells, which are easier to transfect than many primary human cells. In this case , the edited pluripotent stem

cells can be differentiated into monocytes to test for its effect directly on *CARD9* expression.

Assessing for allele specific transcription factors using EMSAs can give us an indication of what may be binding around the variant of interest, but in this case was unsuccessful in refining any allele specific differences in affinity or transcription factors. It may be the case that a different transcription factor may be binding to both alleles, as demonstrated by the ChIP-seq and motif *in silico* predictions. A supershift may have been demonstrated for USF1 and YY1, but for both alleles. The EMSA in primary monocytes did not produce as many complexes as the THP1 EMSA, and suggested additional binding in the A allele. These results should be followed up with ChIP based experiments to determine what may be happening *in vivo* in $CD14^+$ monocytes. The RNA-seq expression data demonstrated no significant difference in *CARD9* expression between AS patients and healthy controls, and also between genotypes, suggesting that *CARD9* is not the functional gene at this locus or that $CD14^+$ monocytes are not relevant.

5.4.4 Gene expression at the 9q32 locus

CARD9 (caspase-associated recruitment domain 9 protein) is highly conserved across species, with a vital role in the innate immune response and therefore has potential relevance in the pathogenesis of AS. eQTLs associated with *CARD9* expression are found in monocytes, suggesting that the innate immune response may be of importance in the aetiology of AS and other SpAs. These SNPs are also associated with mRNA expression of *CARD9* rather than the neighbouring gene *SNAPC4*, which also harbours many AS GWAS SNPs at this locus (Pointon et al. 2010). There has been evidence to suggest a major role for the innate immune response in AS rather than adaptive events as previously thought (De Rycke et al. 2006; Inman 2009; Lories and Baeten 2009). CARD9 is a component

of the NOD- like receptor signaling pathway, a pattern recognition receptor for intracellular bacteria. The NOD2 pathway is involved in bacterial sensing, and mutations downstream may affect the ability to form an effective barrier, causing an uncontrolled inflammatory response.

Associations in *CARD9* are shared between AS and IBD, with CD also being associated with variants in *NOD2* highlighting the potential importance of the NOD2 pathway in these autoinflammatory diseases (Hugot et al. 2001; Ogura et al. 2001). There is known to be overlap between AS and IBD, with 10% of AS patients developing IBD and 70% developing subclinical gut inflammation (Baeten et al. 2002). It was shown that a lack of intestinal inflammation in patients with AS indicates a good prognosis (Mielants et al. 1995), suggesting a link between intestinal health and the development of AS, and a role for the maintenance of gut or microbial homeostasis.

CARD9 also has a role in fine-tuning the anti-microbial response by acting as an adaptor protein linking tyrosine kinases and activation of NF-κB response to TLRs and in the activation of MAPK cascade (Colonna 2007). CARD9 also regulates cytokine activation, such as the IL23 pathway, and up-regulation of the Th17 response and may indicate a link between CARD9 and the IL23 pathway, as polymorphisms in *IL23R* are also associated with AS and IBD (Cortes et al. 2013; Evans et al. 2011; Liu et al. 2015). It is also important to consider environmental influences in triggering disease. The microbiome may be the underlying mechanistic basis for the transition phase leading to chronic inflammation, and it has been demonstrated that HLA-B*27 transgenic mice did not develop joint or intestinal inflammation when raised in the absence of commensal intestinal bacteria (Podolsky 1997). *CARD9* was not found to be significantly differentially expressed in AS patients versus healthy controls in Chapter 3, however, it was shown to be slightly up-regulated in AS patients in $CD14^{+}$ monocytes (FDR=0.096). All other cell types demonstrated a much larger

FDR value, suggesting that $CD14^+$ monocytes are likely to be the most relevant immune cell type in PBMCs relating to the 9q34 association, if *CARD9* is the causal gene at this locus. Our AS patients were not reported to exhibit any gut inflammation, therefore *CARD9* gene expression changes may be more relevant in a subtype of patients that also exhibit overlap with IBD symptoms. The SNPs at this locus may also be more relevant in cells of the intestinal epithelium or colonic mucosa, as these cell types have also been demonstrated as important in AS pathogenesis (Li et al. 2017c; Farh et al. 2015).

SNAPC4 (Small Nuclear RNA Activating Complex Polypeptide 4) is another candidate gene due to a number of associated SNPs that lie within its boundaries. One of these SNPs is a nonsynomymous mutation (rs3812571) that, however, does not appear in the credible set for this locus. *SNAPC4* is involved in RNA polymerase III transcription initiation and gene expression pathways; therefore it may have no relevance in immune function directly. There are also no eQTLs for this gene, suggesting *SNAPC4* does not have a role in AS pathogenesis.

The SNP rs4307445 lies within intron 9 of the *SNAPC4* gene, and was found to be overlapping enhancer ChromHMM marks in GM12878, $CD4^+$ and $CD8^+$ T cells, along with open chromatin in $CD8^+$ ATAC-seq for control samples only. Promoter capture Hi-C data links this SNP to the promoter of *INPP5E* in total activated $CD4^+$ T cells (Javierre et al. 2016). The SNPs at the *CARD9* locus could therefore be exhibiting a cell type specific affect in different SNPs for multiple genes. *INPP5E* (Inositol Polyphosphate-5-Phosphatase E), is involved in mobilising intracellular calcium and acts as a second messenger mediating cell responses to various stimulation. This gene has been implicated in the role of ciliopathies, such as Joubert syndrome (Bielas et al. 2009). An eQTL for *INPP5E* at rs4077515, which is in high LD with rs4307445, is also described in CD and UC in cells from the colon (Hulur et al. 2015), provide further evidence for a plausible role in disease.

Chapter 6

General Discussion

In this thesis I have demonstrated how the integration of multiple functional genomic datasets can begin to piece together the underlying mechanisms involved in AS pathogenesis. This thesis was unable to confirm any new functional mechanisms on the aetiology of AS, but generated many new hypotheses that should be further characterised by functional experiments. This chapter will discuss the findings from this thesis in regards to the larger research aims, and the further work that is required in this endeavour.

6.1 Cell type and context specificity

The expression profile of many genes varies in regards to cell type and changes in the physiological environment (Fairfax et al. 2012; Fairfax et al. 2014). Understanding disease specific changes in expression is therefore necessary in the most relevant cell type. This is not always possible, as often the representative cell type is unknown, or there may be a number of cell types involved with varying functional mechanisms. Studies that have investigated the enrichment of prioritised GWAS SNPs with histone marks representative of enhancer sites found many cell types with potential importance, including monocytes, colonic mucosa, duodenum mucosa, NK cells, small intestine mucosa, sigmoid colon, rectum, regulatory T cells and B cells (Li et al. 2017c; Farh et al. 2015). Due to time constraints and resources, the investigation of only a subset of these cell types in AS was possible, and was restricted to immune cell types $CD14^{+}$

monocytes, $CD16^+$ neutrophils, $CD19^+$ B cells, NK, $CD4^+$ and $CD8^+$ T cells. Across these cell types, differences in significantly differentially expressed genes and pathways were evident, demonstrating the requirement for future tissue specific analyses in understanding disease aetiology. Tissue specificity is also reflected at the chromatin accessibility level, and many large consortium projects have demonstrated the need to profile the epigenetic landscape in a multitude of specific tissue and cell types (Bernstein et al. 2010; Dunham 2012; Adams et al. 2012).

6.2 Transcriptomics in clinical cohorts

Investigating changes in gene expression between diseased and healthy individuals is a common method used in beginning to understand the underlying changes involved in disease pathogenesis. Many of the transcriptomic studies in AS and other immune mediated diseases have focused on using whole blood or PBMCs, which may be one explanation for the lack of replication across studies containing similar groups of patients. This also demonstrates the dangers of post hoc subgroup analyses. This study was able to profile the transcriptional landscape in up to 6 immune cell types, identifying cell specific signatures of genes and pathways that may be relevant in AS. Several genes and pathways, such as NF-κB signalling and *TNFAIP3* were found across innate immune cell types, which may provide increased evidence for their importance. Replication of results is necessary for the prioritisation of future therapeutic targets, along with the requirement to understand the underlying molecular mechanisms that cause these changes in specific groups of patients.

Disease heterogeneity may also be difficult to assess, as patients often differ in relation to disease activity and further disease phenotypes, including peripheral joint involvement, uveitis, and bowel inflammation (Stolwijk et al. 2015a).

Mixing different molecular representations of the disease may reduce the power to detect strong signatures of gene expression if these extra-articular manifestations are also hallmarks of the disease. Individuals within families may present with different SpA subtypes which may suggest a common genetic background for different phenotypic manifestations of the disease (Baeten et al. 2013). Increasing the cohort size with the ability to sub-classify patients, also across SpA subtyptes, may lead to novel findings of general SpA pathogenesis relating to either axial or peripheral disease.

There are a number of cell types believed to be important in AS pathogenesis, such as the small intestine, sigmoid colon and rectum (Li et al. 2017c) in which gene expression changes are yet to be investigated. The gene expression profile of immune cells in AS may also be more inflammatory in the synovial joints of patients than the blood, meaning that detecting disease relevant changes may be more challenging when investigating these immune cells isolated from whole blood. These types of samples are more difficult to obtain as they require more invasive procedures.

6.3 Patient specific chromatin landscape

Assaying the chromatin landscape is able to inform us of transcriptional changes that are currently occurring or will be occurring but are yet to appear at the transcriptional level. They may also describe potential molecular mechanisms regulating gene expression, especially when associated with variants prioritised to be involved with disease, as was explored at the chromosome 2 locus in Chapter 5. This thesis has been able to provide a case for chromatin profiling across different contexts in the same cell type, identifying a number of significant differential sites in $CD14^+$ monocytes related to AS, and in $CD8^+$ T cells comparing anti-TNF treatment against untreated AS patients. This is one of the

first studies to begin to investigate genome-wide chromatin changes in relation to disease, meaning that no generally accepted method to analyse these changes exist. Approaches initially built to analyse gene expression changes were adopted for this analysis. All other comparisons did not reach significance. This may relate to a number of factors, including sample size, analysis method and patient heterogeneity.

Our analysis method was only able to detect differences in the levels of openness at regions pre-defined as accessible in all samples included in the analysis. Therefore it may be the case that a number of differences may be missed between patients and healthy controls; for example, the absence of an ATAC-seq peak, which was described at rs4672505 and potentially at rs4307445 at the 2p15 and *CARD9* loci. There may not be a statistical method to detect these changes unless sample size is substantially increased. Different methods that do not take into account peak calling have been used to detect differences in differential chromatin modification sites using ChIP-seq, and can be used with very low number of replicates (Shen et al. 2013).

6.4 GWAS variant prioritisation

GWAS have provided us with a hypothesis free view of prioritising variants and genes associated with disease across the genome. Often these associations require further refinement, and a number of accepted statistical approaches are used in the attempt to reduce down the number of potentially causal variants, as previously mentioned (Bunt et al. 2015; Consortium et al. 2012). The use of cell type specific annotation has enabled us to define the most relevant cell types for disease (Li et al. 2017c; Farh et al. 2015) along with the functional mechanism of individual regulatory variants, as previously demonstrated at a type 2 diabetes variant that increased enhancer activity in islets (Gaulton et al.

2015). Genotyping data for GWAS cohorts is often not easily accessible, and the fine-mapping analysis presented was only carried out in a small subset of the cohort. Refinement was successful for a large majority of the loci, although varied slightly from the European refinement results (Bunt et al. 2015). Specific functional variants are able to be defined by annotation with functional data, including AS specific ATAC-seq data. This led to the confirmation that rs4672505 at locus 2p15 was most likely the causal SNP at that locus, and is only active in regulating the gene expression of *B3GNT2* in $CD8^+$ T cells.

6.5 Clinical translation

The identification of regulatory variants and differentially expressed genes associated with disease will help to generate new hypotheses underlying disease pathogenesis. The translation of these findings into the clinic remains a huge challenge: for these findings to be of any use, they are required to be robust and therefore validated in further studies.

The use of gene expression assays in the clinic requires the production of a gene signature, that is robust and can be easily measured within hours or minutes. A model to distinguish CD from UC has been reported, which includes a CD specific gene signature, along with the shift in the microbial community in CD. The main gene coexpression signatures characterising the differences between CD and UC was the up-regulation of *DUOX2* and the down-regulation of *APOA1* (Haberman et al. 2014). This type of classification can guide treatment options, as this study also demonstrated that those CD patients with a higher baseline of *APOA1* and increased microbial *Veillonella* abundance relative to *Blautia* were more likely achieve a 6 month remission using anti-TNF therapy (Haberman et al. 2014). This type of finding also suggests that certain gene signatures are able to define groups of patients that are most likely to be responsive to TNF-α

blockers, or further biological therapies that are soon to be used in the clinic. The gene expression analyses presented in this thesis identified many genes that may be prioritised as biomarkers of disease activity and treatment response. These require further validation in larger cohorts and proteomic studies.

ATAC-seq profiling also has potential in the clinic for diagnostic purposes. The time-frame required to carry out an ATAC-seq experiment, excluding cell isolation is approximately half a day (Buenrostro et al. 2013). Targeted approaches using PCR may also keep costs affordable. This approach may be able to stratify patients regarding underlying molecular mechanisms.

6.6 Future work

Continued patient recruitment

The GENExpress ID study is ongoing, and is currently recruiting further AS patients to build numbers in both the transcriptomic and chromatin profiling aspects of the project. AS patients with active disease are eligible; however patient heterogeneity at the molecular level may potentially be affecting our ability to detect strong signatures of disease within these datasets. A solution to overcome this caveat is to increase numbers substantially. This will also enable the sub-classification of patients based on certain gene expression or chromatin changes that may become more evident as power increases. Larger patient numbers will also enable an AS cell type specific eQTL analysis to be conducted. The inclusion of other types of patient specific data across cell types, including ChIP-seq of various chromatin marks and CTCF will also increase knowledge relating to disease specific chromatin changes. Increased power and expanded patient specific datasets will aid in the investigation of disease mechanisms involving enhancers and the regulation of gene expression.

Further exploration of methods for chromatin accessibility analysis

As previously mentioned, this thesis has provided an overview of one method to detect changes in differential open chromatin between AS patients and controls, and also the effect of anti-TNF therapy on chromatin accessibility. The method described assumes that all disease related changes are at defined open chromatin regions in both groups, and that the change in the relative amount of accessible chromatin is related to disease associated changes in gene expression. This method is therefore also reliant on peak calling, discarding regions of the genome that are down-regulated enough not be called as a peak in ATAC-seq. Analysis of this dataset is still ongoing, and the use of programs that do not require peak calling such as diffReps will be investigated (Shen et al. 2013).

Data integration

Data integration is important in defining disease mechanisms, and the addition of further AS specific datasets will enable refinement of disease specific chromatin states, for example by using ChromHMM (Ernst and Kellis 2012). Understanding how GWAS variants may be influencing cell type specific gene expression changes can be identified by utilising eQTL datasets. Colocalisation analysis of eQTLs and GWAS variants would be beneficial in defining regulatory variants associated with disease, along with other datasets such as meQTLs. Defining allele specific expression using the AS RNA-seq data will also be an important addition to understanding the function of regulatory variants in AS.

Functional follow up

Prioritised regulatory variants require functional investigation to prove their mechanism of action in the appropriate cell type. Genomic editing of enhancer regions or disease associated regulatory variants using CRISPR-cas9 can enable functional characterisation in the true cellular context, which plasmid based

approaches such as the functional work described in Chapter 5 does not allow. This approach is currently in progress using induced pluripotent stem cells, with the aim to investigate a number of loci in various immune cell types and their effect on gene expression.

6.7 Concluding remarks

The overall aim of this thesis was to begin to understand the cell type specific expression and chromatin accessibility of immune cell types in AS, and also in response to anti-TNF therapy. These results were also analysed in the context of GWAS, highlighting the importance of integrating multiple data types. The results have highlighted the importance of cell specificity when interpreting disease mechanisms, which will hopefully fuel more targeted future research in AS and SpA.

Chapter 7

Appendices

7.1 Additional Tables

7.1.1 Chapter 3 Tables

Table 7.1: Patient metadata including individual BASDAI criteria. Scores relating to pain and severity answered by the patient, with 0 meaning no pain and 10 meaning the worst pain. Blood test markers for inflammation; CRP= C-reactive protein (mg/L), ESR= Erythocyte sedimentation rate (mm/hr). *High sensitivity CRP test for 1 patient due to low CRP levels.

Patient ID	Fatigue	Spinal pain	Peripheral joint	Tenderness	Stiffness	Stiffness duration	CRP(mg/L)	ESR(mm)
GNORLAS004	7.4	7.8	8.1	7.1	9.2	8.3	90.7	54
GNORLAS014	8.0	4.0	5.0	4.0	7.0	5.0	0.28	-
GNORLAS017	6.4	6.5	0.6	1.8	7.4	7.5	-	-
GNORLAS018	-	-	-	-	-	-	11.2	8.0
GNORLAS019	7.5	7.5	7.5	7.5	7.5	7.5	0.074mg/dl*	-
GNORLAS035	5.8	7.6	5.0	6.7	7.1	4.3	13.4	7.0
GNORLAS036	7.4	0.3	2.6	0.3	4.7	1.2	-	-
GNORLAS037	1.3	0.3	0.0	0.0	1.1	2.3	1.4	-
GNORLAS038	4.8	3.6	0.0	0.0	3.7	1.2	3.2	2
GNORLAS041	0.0	2.0	0.0	1.2	2.0	2.5	-	-

7.1.2 Chapter 4 Tables

Table 7.2: Patient metadata including individual BASDAI criteria before beginning on biologic therapy. Scores relating to pain and severity answered by the patient, with 0 meaning no pain and 10 meaning the worst pain. Blood test markers for inflammation; CRP= C-reactive protein (mg/L), ESR= Erythocyte sedimentation rate (mm/hr).

Patient ID	Fatigue	Spinal pain	Peripheral joint	Tenderness	Stiffness	Stiffness duration	CRP(mg/L)	ESR(mm)
GNOPBAS043	8.6	8.6	8.6	7.8	9.6	10.0	2.7	6.0
GNOPBAS044	7.4	6.6	6.5	6.8	7.8	10.0	-	-
GNOPBAS045	5.4	5	7.5	3.7	4.6	7.3	30.2	18
GNOPBAS046	8.0	7.5	7.5	7.3	7.6	7.5	7.2	9.0
GNOPBAS047	8.8	8	7.6	1.6	6.2	4.8	8.8	6.0

Table 7.3: Patient metadata including individual BASDAI criteria post 3 months biologic therapy. Scores relating to pain and severity answered by the patient, with 0 meaning no pain and 10 meaning the worst pain. Blood test markers for inflammation; CRP= C-reactive protein (mg/L), ESR= Erythocyte sedimentation rate (mm/hr).

Patient ID	Fatigue	Spinal pain	Peripheral joint	Tenderness	Stiffness	Stiffness duration	CRP(mg/L)	ESR(mm)
GNOPBAS043	7.7	7.7	6.4	5.2	8.2	2	2.2	8
GNOPBAS044	4.3	3.3 2.0	1.0	1.8	4.7	-	2	
GNOPBAS045	0.2	0.05	0.1	0.2	0.1	0.2	6.0	5.0
GNOPBAS046	4.2	3.3	2.9	3.8	4.2	5.0	0.7	2.0

Enriched pathways $CD14^+$ monocytes FDR <0.01
CARM1 and Regulation of the Estrogen Receptor
Mechanism of Gene Regulation by Peroxisome Proliferators via PPAR-α
HIF-2-alpha transcription factor network
Telomeres, Telomerase, Cellular Aging, and Immortality
IL2 signaling events mediated by PI3K
Notch-mediated HES/HEY network
Complement cascade
Glypican 1 network
Signaling events mediated by HDAC Class III
NOTCH1 Intracellular Domain Regulates Transcription
Signaling by NOTCH1
Granule Cell Survival Pathway is a specific case of more general PAC1 Receptor Pathway
Signaling by NOTCH
Complement and coagulation cascades
Transport to the Golgi and subsequent modification
C-MYB transcription factor network

Table 7.4: Pathways enriched in $CD14^+$ monocytes between AS patients and their paired 3 month follow up sample following biologics treatment with TNF-α inhibitors, using differentially expressed genes FDR <0.01.

Enriched pathways $CD4^+$ T cells FDR <0.01
GPCR downstream signaling
Genes encoding secreted soluble factors
Cytokine-cytokine receptor interaction

Table 7.5: Pathways enriched in $CD4^+$ T cells between AS patients and their paired 3 month follow up sample following biologics treatment with TNF-α inhibitors, using differentially expressed genes FDR <0.05 (the number of DE genes at FDR <0.01 is too small).

Enriched pathways $CD8^+$ T cells FDR <0.01
Pertussis toxin-insensitive CCR5 Signaling in Macrophage
HIV-1 Nef: Negative effector of Fas and TNF-alpha
Chemokine receptors bind chemokines
Leucine rich repeat containing receptor (NLR) signaling pathways
Canonical NF-kappaB pathway
Keratinocyte Differentiation
Cytokine-cytokine receptor interaction
G alpha q Pathway
Downstream signaling in naive $CD8^+$ T cells
IL2-mediated signaling events
IL2 signaling events mediated by STAT5
RIG-I/MDA5 mediated induction of IFN-α/β pathways
GPCR downstream signaling
Signaling by Rho GTPases
T Cell Receptor Signaling Pathway
Chemokine signaling pathway
Mechanism of Gene Regulation by Peroxisome Proliferators via PPAR-α
Thromboxane A2 receptor signaling
ATF-2 transcription factor network
NOD-like receptor signaling pathway
Validated targets of C-MYC transcriptional repression
LPA receptor mediated events
IL12-mediated signaling events
Innate Immune System
p75 NTR receptor-mediated signalling

Table 7.6: Pathways enriched in $CD8^+$T cells between AS patients and their paired 3 month follow up sample following biologics treatment with TNF-α inhibitors, using differentially expressed genes FDR <0.01.

7.1.3 Chapter 5 Tables

Locus	Description
1q23 (*FCGR2A*)	Bayes factor does not reach significance
1q26 (*RUNX3*)	Bayes factor does not reach significance
2q31 (*UBE2E3*)	Unable to find association
5q33 (*IL12B*)	Bayes factor does not reach significance
6q15 (*BACH2*)	Unable to find association
12q24 (*SH2B3*)	Unable to find association
16p12 (*IL27*)	Unable to find association
17q21 (*TBKBP1*)	Bayes factor does not reach significance
19p13 (*TYK2*)	Unable to find association
21q22 (*ICOSLG*)	Unable to find association

Table 7.7: Fine mapping of Immunochip AS associated loci that did not reach significance. 11 AS associated loci were either unable to be fine mapped, or did not reach a significance threshold of a log10 approximate bayes factor of 5.

7.2 Additional Figures

7.2.1 Chapter 3 Figures

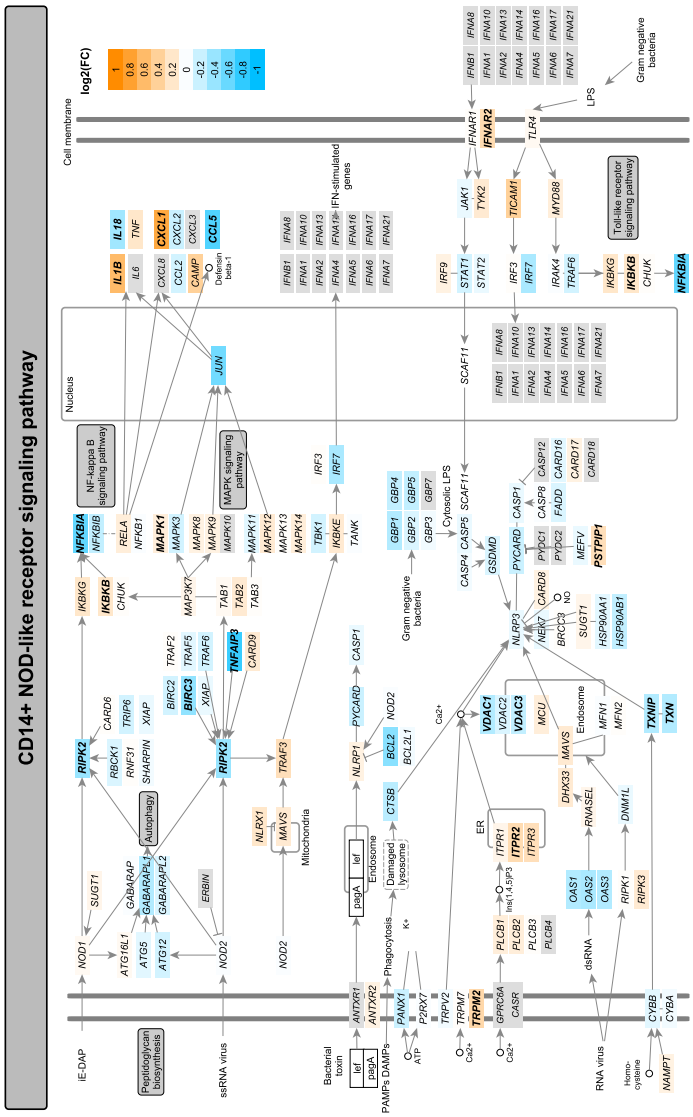
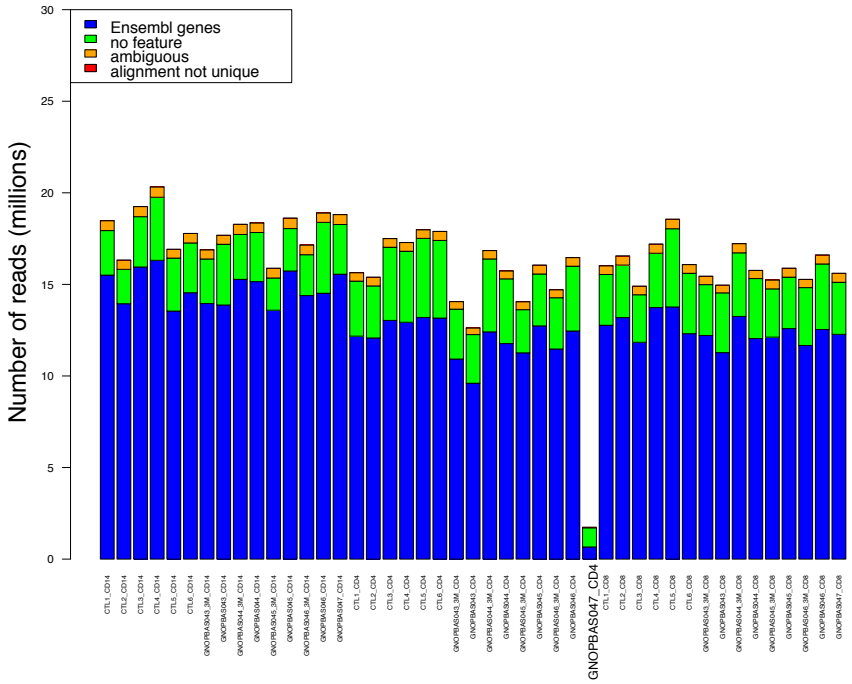
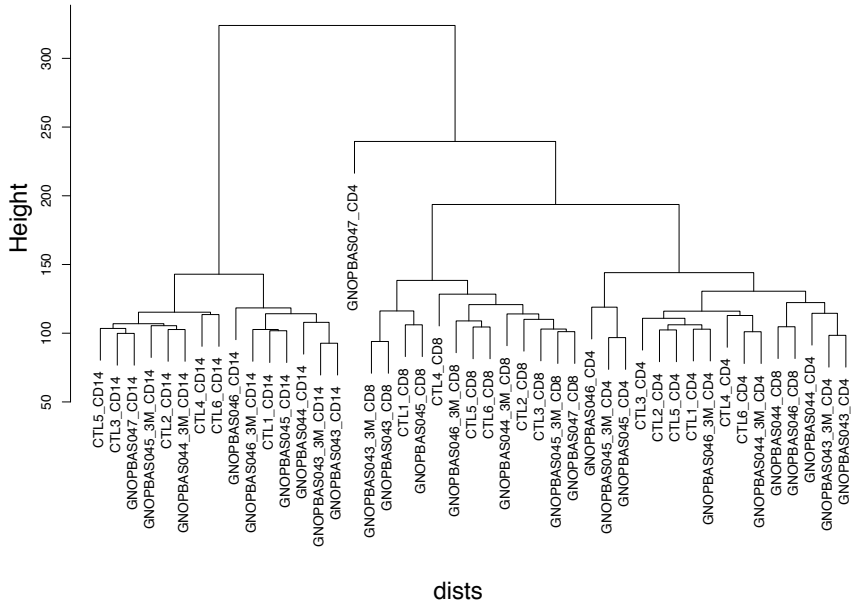


Figure 7.1: The NOD-like receptor signalling pathway in CD14$^+$ monocytes demonstrating the gene expression changes in AS patients versus healthy controls. Genes are coloured by their log2 Fold Change, and significant genes are in bold.

7.2.2 Chapter 4 Figures



(a)



(b)

Figure 7.2: Quality Control of AS cohort 2 RNA-seq data. **(a)** Read assignment of the all samples mapping to known mRNA and lncRNA genes, **(b)** Unsupervised hierarchical clustering, demonstrating clear seperation between cell types and GNOPBAS047 $CD4^{+}$.

7.2.3 Chapter 5 Figures

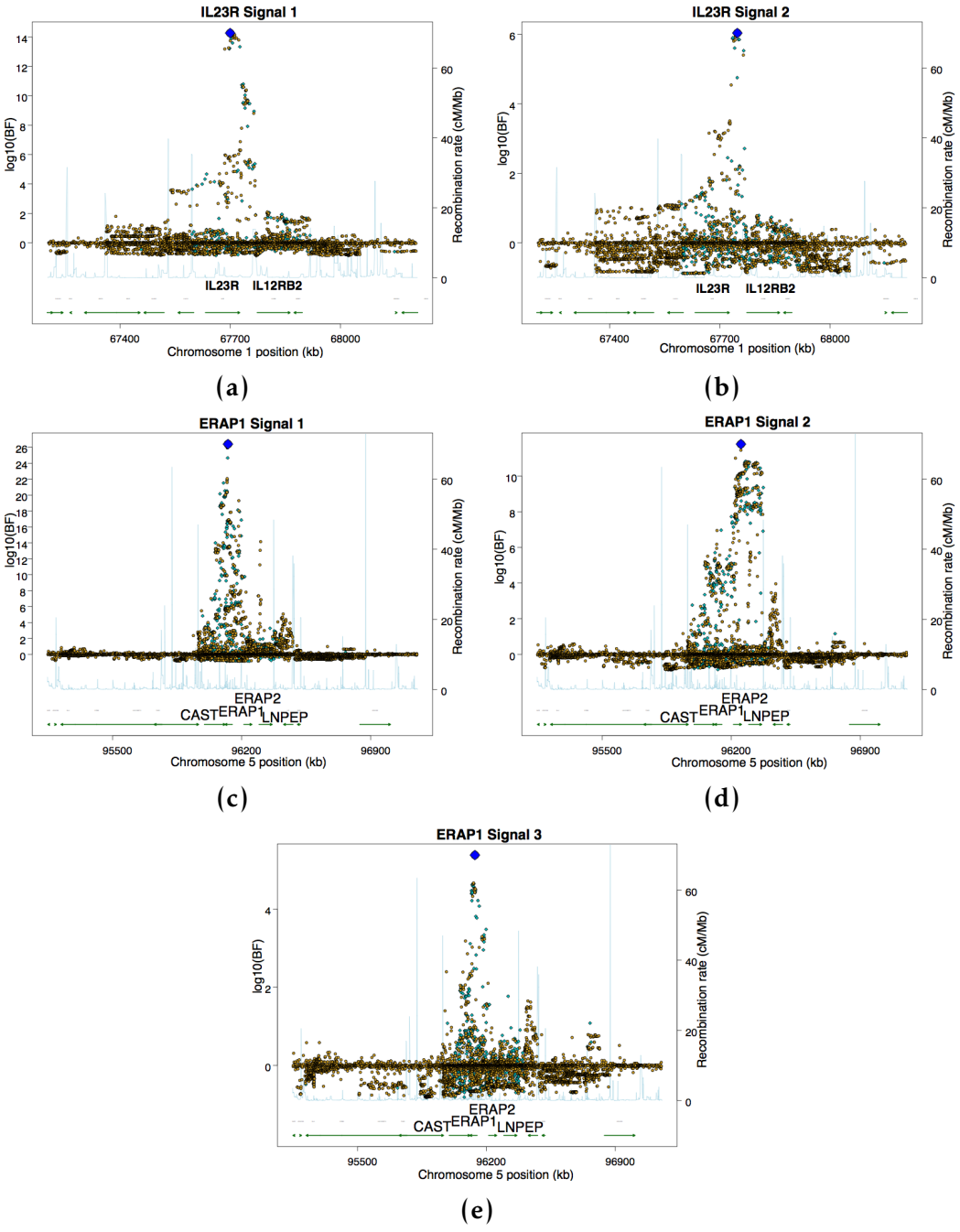
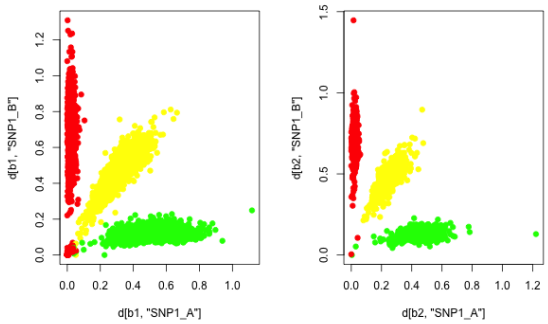
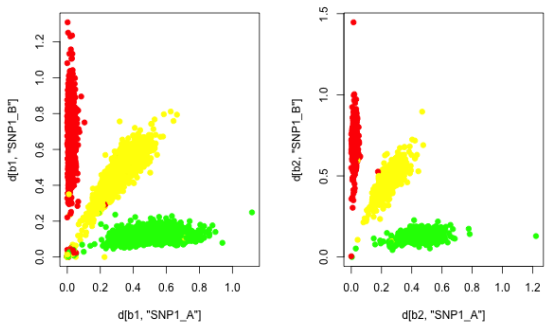


Figure 7.3: Local association plots of the log10 approximate bayes factor demonstrating multiple association signals at a single locus. at **(a)** *IL23R* **(b)** *IL23R* conditional **(c)** *ERAP1* **(d)** *ERAP1* conditional **(e)** *ERAP1* signal conditioning on the first two signals. Recombination hotspots are also demonstrated in blue, and imputed SNPs are shown in gold.



(a) rs1128905 Genotyped



(b) rs1128905 Imputed

Figure 7.4: rs1128905 is called in a different genotype cluster following imputation. This demonstrates an error in genotyping between cases and controls in the original GWAS analysis, as shown by the red point appearing in the middle cluster after imputation. This analysis was carried out by Dr Adrián Cortés with the original IGAS Immunochip cohort.

Bibliography

Abdollahi, E., Tavasolian, F., Momtazi-Borojeni, A. A., Samadi, M., and Rafatpanah, H. (2016). "Protective role of R381Q (rs11209026) polymorphism in IL-23R gene in immune-mediated diseases: A comprehensive review". *J Immunotoxicol* 13: pp. 286–300. ISSN: 1547-691x. DOI: 10.3109/1547691x.2015.1115448.

Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T., and McVean, G. A. (2012). "An integrated map of genetic variation from 1,092 human genomes". *Nature* 491: pp. 56–65. ISSN: 0028-0836. DOI: 10.1038/nature11632.

Abraham, C. and Cho, J. (2009). "Interleukin-23/Th17 pathways and inflammatory bowel disease". *Inflamm Bowel Dis* 15: pp. 1090–100. ISSN: 1078-0998. DOI: 10.1002/ibd.20894.

Adams, D., Altucci, L., Antonarakis, S. E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A., et al. (2012). "BLUEPRINT to decode the epigenetic signature written in blood". *Nat Biotechnol*. Vol. 30. United States: pp. 224–6. ISBN: 1546-1696 (Electronic) 1087-0156 (Linking). DOI: 10.1038/nbt.2153.

Adcock, I. M. (2006). "Histone deacetylase inhibitors as novel anti-inflammatory agents". *Curr Opin Investig Drugs* 7: pp. 966–73. ISSN: 1472-4472 (Print) 1472-4472.

Adrianto, I., Wen, F., Templeton, A., Wiley, G., King, J. B., Lessard, C. J., Bates, J. S., Hu, Y., Kelly, J. A., Kaufman, K. M., et al. (2011). "Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus". *Nat Genet* 43: pp. 253–8. ISSN: 1061-4036. DOI: 10.1038/ng.766.

Ahern, P. P., Schiering, C., Buonocore, S., McGeachy, M. J., Cua, D. J., Maloy, K. J., and Powrie, F. (2010). "Interleukin-23 drives intestinal inflammation through direct activity on T cells". *Immunity* 33: pp. 279–88. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2010.08.010.

Alvarez-Garcia, I. and Miska, E. A. (2005). "MicroRNA functions in animal development and human disease". *Development* 132: pp. 4653–62. ISSN: 0950-1991 (Print) 0950-1991. DOI: 10.1242/dev.02073.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., Dreyfuss, G., Eddy, S. R., Griffiths-Jones, S., Marshall, M., et al. (2003). "A uniform system for microRNA annotation". *Rna* 9: pp. 277–9. ISSN: 1355-8382 (Print) 1355-8382.

Anandarajah, A. and Ritchlin, C. T. (2005). "Treatment update on spondyloarthropathy". *Curr Opin Rheumatol* 17: pp. 247–56. ISSN: 1040-8711 (Print) 1040-8711.

Anders, S., Pyl, P. T., and Huber, W. (2014). "HTSeqa Python framework to work with high-throughput sequencing data". *HTSeqa Python framework to work with high-throughput sequencing data*. DOI: 10.1093/bioinformatics/btu638. URL: http://dx.doi.org/10.1093/bioinformatics/btu638.

Anderson, C. A., Boucher, G., Lees, C. W., Franke, A., D'Amato, M., Taylor, K. D., Lee, J. C., Goyette, P., Imielinski, M., Latiano, A., et al. (2011). "Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47". *Nat Genet* 43: pp. 246–52. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.764.

Arcaroli, J., Silva, E., Maloney, J. P., He, Q., Svetkauskaite, D., Murphy, J. R., and Abraham, E. (2006). "Variant IRAK-1 haplotype is associated with increased nuclear factor-kappaB activation and worse outcomes in sepsis". *Am J Respir Crit Care Med* 173: pp. 1335–41. ISSN: 1073-449X (Print) 1073-449x. DOI: 10.1164/rccm.200603-341OC.

Arends, S., Brouwer, E., Veer, E. van der, Groen, H., Leijsma, M. K., Houtman, P. M., Th, A. Jansen T. L., Kallenberg, C. G., and Spoorenberg, A. (2011). "Baseline predictors of response and discontinuation of tumor necrosis factor-alpha blocking therapy in ankylosing spondylitis: a prospective longitudinal observational cohort study". *Arthritis Res Ther* 13: R94. ISSN: 1478-6354. DOI: 10.1186/ar3369.

Armas, J. B., Gonzalez, S., Martinez-Borra, J., Laranjeira, F., Ribeiro, E., Correia, J., Ferreira, M. L., Toste, M., Lopez-Vazquez, A., and Lopez-Larrea, C. (1999). "Susceptibility to ankylosing spondylitis is independent of the Bw4 and Bw6 epitopes of HLA-B27 alleles". *Tissue Antigens* 53: pp. 237–43. ISSN: 0001-2815 (Print) 0001-2815.

Arrieta, M. C., Bistritz, L., and Meddings, J. B. (2006). "Alterations in intestinal permeability". *Gut* 55: pp. 1512–1520. ISSN: 0017-5749 1468-3288. DOI: 10.1136/gut.2005.085373. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1856434/.

Askin, A., Guvendi, E., Tosun, A., and Demirdal, U. S. (2017). "Paradoxical Side Effect Related With Anti-Tumor Necrosis Factor Alpha Treatment". *Med Arch* 71: pp. 148–150. ISSN: 0350-199X (Print) 0350-199x. DOI: 10.5455/medarh.2017.71.148-150.

Assassi, S., Reveille, J. D., Arnett, F. C., Weisman, M. H., Ward, M. M., Agarwal, S. K., Gourh, P., Bhula, J., Sharif, R., Sampat, K., et al. (2011). "Whole-blood Gene Expression Profiling in Ankylosing Spondylitis Shows Upregulation of Toll-like Receptor 4 and 5". *J Rheumatol* 38: pp. 87–98. ISSN: 0315-162X (Print). DOI: 10.3899/jrheum.100469.

Atianand, M. K. and Fitzgerald, K. A. (2014). "Long non-coding RNAs and control of gene expression in the immune system". *Trends Mol Med* 20: pp. 623–31. ISSN: 1471-4914. DOI: 10.1016/j.molmed.2014.09.002.

Atteno, M., Peluso, R., Costa, L., Padula, S., Iervolino, S., Caso, F., Sanduzzi, A., Lubrano, E., Del Puente, A., and Scarpa, R. (2010). "Comparison of effectiveness and safety of infliximab, etanercept, and adalimumab in psoriatic arthritis patients who experienced an inadequate response to previous disease-modifying antirheumatic drugs". *Clin Rheumatol* 29: pp. 399–403. ISSN: 0770-3198. DOI: 10.1007/s10067-009-1340-7.

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., and Abecasis, G. R. (2015). "A global reference for human genetic variation". *Nature* 526: pp. 68–74. ISSN: 0028-0836. DOI: 10.1038/nature15393.

Babon, J. J., Kershaw, N. J., Murphy, J. M., Varghese, L. N., Laktyushin, A., Young, S. N., Lucet, I. S., Norton, R. S., and Nicola, N. A. (2012). "Suppression of cytokine signaling by SOCS3: characterization of the mode of inhibition and the basis of its specificity". *Immunity* 36: pp. 239–50. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2011.12.015.

Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). "The impact of microRNAs on protein output". *Nature* 455: pp. 64–71. ISSN: 0028-0836. DOI: 10.1038/nature07242.

Baeten, Dominique, De Keyser, Filip, Mielants, Herman, and Veys, Eric M. (2002). "Ankylosing spondylitis and bowel disease". *Best Practice & Research Clinical Rheumatology* 16: pp. 537–549. ISSN: 1521-6942. DOI: http://dx.doi.org/10.1053/berh.2002.0249. URL: http://www.sciencedirect.com/science/article/pii/S1521694202902494.

Baeten, Dominique, Breban, Maxime, Lories, Rik, Schett, Georg, and Sieper, Joachim (2013). "Are spondylarthritides related but distinct conditions or a single disease with a heterogeneous phenotype?" *Arthritis & Rheumatism* 65: pp. 12–20. ISSN: 1529-0131. DOI: 10.1002/art.37829. URL: http://dx.doi.org/10.1002/art.37829.

Banovich, N. E., Lan, X., McVicker, G., Geijn, B. van de, Degner, J. F., Blischak, J. D., Roux, J., Pritchard, J. K., and Gilad, Y. (2014). "Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels". *PLoS Genet* 10: e1004663. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1004663.

Baraliakos, X., Berg, R. van den, Braun, J., and Heijde, D. van der (2012). "Update of the literature review on treatment with biologics as a basis for the first update of the ASAS/EULAR management recommendations of ankylosing spondylitis". *Rheumatology (Oxford)* 51: pp. 1378–87. ISSN: 1462-0324. DOI: 10.1093/rheumatology/kes026.

Barnes, P. J. (2006). "How corticosteroids control inflammation: Quintiles Prize Lecture 2005". *Br J Pharmacol* 148: pp. 245–54. ISSN: 0007-1188 (Print) 0007-1188. DOI: 10.1038/sj.bjp.0706736.

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P., and Marth, G. T. (2011). "BamTools: a C++ API and toolkit for analyzing and managing BAM files". *Bioinformatics* 27: pp. 1691–2. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr174.

Bautista-Caro, Mara-Beln, Arroyo-Villa, Irene, Castillo-Gallego, Concepcin, Miguel, Eugenio de, Peiteado, Diana, Puig-Krger, Amaya, Martn-Mola, Emilio, and Miranda-Cars, Mara-Eugenia (2013). "Decreased Th17 and Th1 cells in the peripheral blood of patients with early non-radiographic axial spondyloarthritis: a marker of disease activity in HLA-B27+ patients". *Rheumatology* 52. ISSN: 1462-0324. DOI: 10.1093/rheumatology/kes267. URL: http://dx.doi.org/10.1093/rheumatology/kes267.

Bei, L., Lu, Y., Bellis, S. L., Zhou, W., Horvath, E., and Eklund, E. A. (2007). "Identification of a HoxA10 activation domain necessary for transcription of the gene encoding beta3 integrin during myeloid differentiation". *J Biol Chem* 282: pp. 16846–59. ISSN: 0021-9258 (Print) 0021-9258. DOI: 10.1074/jbc.M609744200.

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). "Hi-C: a comprehensive technique to capture the conformation of genomes". *Methods* 58: pp. 268–76. ISSN: 1046-2023. DOI: 10.1016/j.ymeth.2012.05.001.

Benjamin, R. and Parham, P. (1990). "Guilt by association: HLA-B27 and ankylosing spondylitis". *Immunol Today* 11: pp. 137–42. ISSN: 0167-5699 (Print) 0167-5699.

Benjamin, R. and Parham, P. (1992). "HLA-B27 and disease: a consequence of inadvertent antigen presentation?" *Rheum Dis Clin North Am* 18: pp. 11–21. ISSN: 0889-857X (Print) 0889-857x.

Beringer, Audrey, Noack, Melissa, and Miossec, Pierre (2016). "IL-17 in Chronic Inflammation: From Discovery to Targeting". *Trends in molecular medicine* 22: pp. 230–241. ISSN: 1471-4914. DOI: 10.1016/j.molmed.2016.01.001. URL: http://dx.doi.org/10.1016/j.molmed.2016.01.001.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., et al. (2010). "The NIH Roadmap Epigenomics Mapping Consortium". *Nat Biotechnol* 28: pp. 1045–8. ISSN: 1087-0156. DOI: 10.1038/nbt1010-1045.

Bielas, S. L., Silhavy, J. L., Brancati, F., Kisseleva, M. V., Al-Gazali, L., Sztriha, L., Bayoumi, R. A., Zaki, M. S., Abdel-Aleem, A., Rosti, R. O., et al. (2009). "Mutations in INPP5E, encoding inositol polyphosphate-5-phosphatase E, link phosphatidyl inositol signaling to the ciliopathies". *Nat Genet* 41: pp. 1032–6. ISSN: 1061-4036. DOI: 10.1038/ng.423.

Blanco-Gomez, A., Castillo-Lluva, S., Del Mar Saez-Freire, M., Hontecillas-Prieto, L., Mao, J. H., Castellanos-Martin, A., and Perez-Losada, J. (2016). "Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes". *Bioessays* 38: pp. 664–73. ISSN: 0265-9247. DOI: 10.1002/bies.201600084.

Bogunia-Kubik, K., Wysoczanska, B., Piatek, D., Iwaszko, M., Ciechomska, M., and Swierkot, J. (2016). "Significance of Polymorphism and Expression of miR-146a and NFkB1 Genetic Variants in Patients with Rheumatoid Arthritis". *Arch Immunol Ther Exp (Warsz)* 64: pp. 131–136. ISSN: 0004-069x. DOI: 10.1007/s00005-016-0443-5.

Bohmdorfer, G. and Wierzbicki, A. T. (2015). "Control of Chromatin Structure by Long Noncoding RNA". *Trends Cell Biol* 25: pp. 623–32. ISSN: 0962-8924. DOI: 10.1016/j.tcb.2015.07.002.

Bonder, Marc J., Luijk, Ren, Zhernakova, Daria V., Moed, Matthijs, Deelen, Patrick, Vermaat, Martijn, Iterson, Maarten van, Dijk, Freerk van, Galen, Michiel van, Bot, Jan, et al. (2017). "Disease variants alter transcription factor levels and methylation of their binding sites". *Nature genetics* 49: pp. 131–138. ISSN: 1061-4036. DOI: 10.1038/ng.3721. URL: http://dx.doi.org/10.1038/ng.3721.

Boonen, A. and Severens, J. L. (2002). "Ankylosing spondylitis: what is the cost to society, and can it be reduced?" *Best Pract Res Clin Rheumatol* 16: pp. 691–705. ISSN: 1521-6942 (Print) 1521-6942.

Borghesi, L. and Milcarek, C. (2007). "Innate versus adaptive immunity: a paradigm past its prime?" *Cancer Res* 67: pp. 3989–93. ISSN: 0008-5472 (Print) 0008-5472. DOI: 10.1158/0008-5472.can-07-0182.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., et al. (2004). "A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway". *Nat Cell Biol* 6: pp. 97–105. ISSN: 1465-7392 (Print) 1465-7392. DOI: 10.1038/ncb1086.

Bowness, P. (2015). "HLA-B27". *Annu Rev Immunol* 33: pp. 29–48. ISSN: 0732-0582. DOI: 10.1146/annurev-immunol-032414-112110.

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., et al. (2012). "Annotation of functional variation in personal genomes using RegulomeDB". *Genome Res* 22: pp. 1790–7. ISSN: 1088-9051. DOI: 10.1101/gr.137323.112.

Bradley, S. J., Suarez-Fueyo, A., Moss, D. R., Kyttaris, V. C., and Tsokos, G. C. (2015). "T Cell Transcriptomes Describe Patient Subtypes in Systemic Lupus Erythematosus". *PLoS One* 10: e0141171. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0141171.

Brandtzaeg, P. (2007). "Induction of secretory immunity and memory at mucosal surfaces". *Vaccine* 25: pp. 5467–84. ISSN: 0264-410X (Print) 0264-410x. DOI: 10.1016/j.vaccine.2006.12.001.

Braun, J., Golder, W., Bollow, M., Sieper, J., and Heijde, D. van der (2002). "Imaging and scoring in ankylosing spondylitis". *Clin Exp Rheumatol* 20: S178–84. ISSN: 0392-856X (Print) 0392-856x.

Braun, J., Berg, R. van den, Baraliakos, X., Boehm, H., Burgos-Vargas, R., Collantes-Estevez, E., Dagfinrud, H., Dijkmans, B., Dougados, M., Emery, P., et al. (2011). "2010 update of the ASAS/EULAR recommendations for the management of ankylosing spondylitis". *Ann Rheum Dis* 70: pp. 896–904. ISSN: 0003-4967. DOI: 10.1136/ard.2011.151027.

Brenowitz, M., Senear, D. F., and Kingston, R. E. (2001). "DNase I footprint analysis of protein-DNA binding". *Curr Protoc Mol Biol* Chapter 12: Unit 12.4. ISSN: 1934-3647. DOI: 10.1002/0471142727.mb1204s07.

Brewerton, D. A., Hart, F. D., Nicholls, Anne, Caffrey, Maeve, James, D. C. O., and Sturrock, R. D. (1973). "ANKYLOSING SPONDYLITIS AND HL-A 27". *The Lancet* 301: pp. 904–907. ISSN: 0140-6736. DOI: http://dx.doi.org/10.1016/S0140-6736(73)91360-3. URL: http://www.sciencedirect.com/science/article/pii/S0140673673913603.

Brown, M., Jepson, A., Young, A., Whittle, H., Greenwood, B., and Wordsworth, B. (1997). "Ankylosing spondylitis in West Africansevidence for a non-HLA-B27 protective effect". *Annals of the Rheumatic Diseases* 56: pp. 68–70. ISSN: 0003-4967 1468-2060. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1752251/.

Brown, M. A., Edwards, S., Hoyle, E., Campbell, S., Laval, S., Daly, A. K., Pile, K. D., Calin, A., Ebringer, A., Weeks, D. E., et al. (2000). "Polymorphisms of the CYP2D6 gene increase susceptibility to ankylosing spondylitis". *Hum Mol Genet* 9: pp. 1563–6. ISSN: 0964-6906 (Print) 0964-6906 (Linking).

Brown, M. A., Crane, A. M., and Wordsworth, B. P. (2002). "Genetic aspects of susceptibility, severity, and clinical expression in ankylosing spondylitis". *Curr Opin Rheumatol* 14: pp. 354–60. ISSN: 1040-8711 (Print) 1040-8711 (Linking).

Brown, Matthew A., Kenna, Tony, and Wordsworth, B. P. (2015). "Genetics of ankylosing spondylitis-insights into pathogenesis". *Nature reviews. Rheumatology*. ISSN: 1759-4790. DOI: 10.1038/nrrheum.2015.133. URL: http://dx.doi.org/10.1038/nrrheum.2015.133.

Buenrostro, Jason D., Giresi, Paul G., Zaba, Lisa C., Chang, Howard Y., and Greenleaf, William J. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". *Nature methods* 10: pp. 1213–1218. ISSN: 1548-7091. DOI: 10.1038/nmeth.2688. URL: http://dx.doi.org/10.1038/nmeth.2688.

Buenrostro, Jason D., Wu, Beijing, Chang, Howard Y., and Greenleaf, William J. (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide". *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* 109: p. 9. ISSN: 1934-3647. DOI: 10.1002/0471142727.mb2129s109. URL: http://dx.doi.org/10.1002/0471142727.mb2129s109.

Bunt, Martijn van de, Cortes, Adrian, Consortium, Igas, Brown, Matthew A., Morris, Andrew P., and McCarthy, Mark I. (2015). "Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci". *PLoS genetics* 11. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1005535. URL: http://dx.doi.org/10.1371/journal.pgen.1005535.

Burton, P R; WTCCC (2007). "Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants". *Nat Genet* 39: pp. 1329–1337. ISSN: 1061-4036. DOI: http://www.nature.com/ng/journal/v39/n11/suppinfo/ng.2007.17_S1.html. URL: http://dx.doi.org/10.1038/ng.2007.17.

Caballero-Franco, C. and Kissler, S. (2016). "The autoimmunity-associated gene RGS1 affects the frequency of T follicular helper cells". *Genes Immun* 17: pp. 228–38. ISSN: 1466-4879. DOI: 10.1038/gene.2016.16.

Calao, M., Burny, A., Quivy, V., Dekoninck, A., and Van Lint, C. (2008). "A pervasive role of histone acetyltransferases and deacetylases in an NF-kappaB-signaling code". *Trends Biochem Sci* 33: pp. 339–49. ISSN: 0968-0004 (Print) 0968-0004. DOI: 10.1016/j.tibs.2008.04.015.

Care, A., Catalucci, D., Felicetti, F., Bonci, D., Addario, A., Gallo, P., Bang, M. L., Segnalini, P., Gu, Y., Dalton, N. D., et al. (2007). "MicroRNA-133 controls cardiac hypertrophy". *Nat Med* 13: pp. 613–8. ISSN: 1078-8956 (Print) 1078-8956. DOI: 10.1038/nm1582.

Carithers, L. J. and Moore, H. M. (2015). "The Genotype-Tissue Expression (GTEx) Project". *Biopreserv Biobank* 13: pp. 307–8. ISSN: 1947-5543. DOI: 10.1089/bio.2015.29031.hmm.

Carpenter, S., Aiello, D., Atianand, M. K., Ricci, E. P., Gandhi, P., Hall, L. L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J. B., et al. (2013). "A long noncoding RNA mediates both activation and repression of immune response genes". *Science* 341: pp. 789–92. ISSN: 0036-8075. DOI: 10.1126/science.1240925.

Carter, K. W., Pluzhnikov, A., Timms, A. E., Miceli-Richard, C., Bourgain, C., Wordsworth, B. P., Jean-Pierre, H., Cox, N. J., Palmer, L. J., Breban, M., et al. (2007). "Combined analysis of three whole genome linkage scans for Ankylosing Spondylitis". *Rheumatology* 46: pp. 763–771. ISSN: 1462-0324. DOI: 10.1093/rheumatology/kel443. URL: http://dx.doi.org/10.1093/rheumatology/kel443.

Chai, W., Lian, Z., Chen, C., Liu, J., Shi, L. L., and Wang, Y. (2013). "JARID1A, JMY, and PTGER4 polymorphisms are related to ankylosing spondylitis in Chinese Han patients: a case-control study". *PLoS One* 8: e74794. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0074794.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets". *Gigascience* 4: p. 7. ISSN: 2047-217X. DOI: 10.1186/s13742-015-0047-8.

Chang, Y. T., Chou, C. T., Shiao, Y. M., Lin, M. W., Yu, C. W., Chen, C. C., Huang, C. H., Lee, D. D., Liu, H. N., Wang, W. J., et al. (2006). "Psoriasis vulgaris in Chinese individuals is associated with PSORS1C3 and CDSN genes". *Br J Dermatol* 155: pp. 663–9. ISSN: 0007-0963 (Print) 0007-0963. DOI: 10.1111/j.1365-2133.2006.07420.x.

Chatzikyriakidou, A., Voulgari, P. V., Georgiou, I., and Drosos, A. A. (2010). "The role of microRNA-146a (miR-146a) and its target IL-1R-associated kinase (IRAK1) in psoriatic arthritis susceptibility". *Scand J Immunol* 71: pp. 382–5. ISSN: 0300-9475. DOI: 10.1111/j.1365-3083.2010.02381.x.

Chen, K., Wei, X. Z., Zhu, X. D., Bai, Y. S., Chen, Y., Wang, C. F., Chen, Z. Q., and Li, M. (2013). "Whole-blood gene expression profiling in ankylosing spondylitis identifies novel candidate genes that may contribute to the inflammatory and tissue-destructive disease aspects". *Cell Immunol* 286: pp. 59–64. ISSN: 0008-8749. DOI: 10.1016/j.cellimm.2013.10.009.

Chen, Xingqi, Shen, Ying, Draper, Will, Buenrostro, Jason D., Litzenburger, Ulrike, Cho, Seung W., Satpathy, Ansuman T., Carter, Ava C., Ghosh, Rajarshi P., East-Seletsky, Alexandra, et al. (2016). "ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing". *Nature methods*. ISSN: 1548-7091. DOI: 10.1038/nmeth.4031. URL: http://dx.doi.org/10.1038/nmeth.4031.

Cheng, Christine S, Gate, Rachel E, Aiden, Aviva P, Siba, Atsede, Tabaka, Marcin, Lituiev, Dmytro, Machol, Ido, Subramaniam, Meena, Shammim, Muhammad, Hougen, Kendrick L, et al. (2016). "Genetic determinants of chromatin accessibility and gene regulation in T cell activation across human individuals". *bioRxiv*. DOI: 10.1101/090241.

Chou, C. H., Chang, N. W., Shrestha, S., Hsu, S. D., Lin, Y. L., Lee, W. H., Yang, C. D., Hong, H. C., Wei, T. Y., Tu, S. J., et al. (2016). "miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database". *Nucleic Acids Res* 44: pp. D239–47. ISSN: 0305-1048. DOI: 10.1093/nar/gkv1258.

Chow, C. Y., Wang, X., Riccardi, D., Wolfner, M. F., and Clark, A. G. (2015). "The genetic architecture of the genome-wide transcriptional response to ER stress in the mouse". *PLoS Genet* 11: e1004924. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1004924.

Chu, Y., Vahl, J. C., Kumar, D., Heger, K., Bertossi, A., Wojtowicz, E., Soberon, V., Schenten, D., Mack, B., Reutelshofer, M., et al. (2011). "B cells lacking the tumor suppressor TNFAIP3/A20 display impaired differentiation and hyperactivation and cause inflammation and autoimmunity in aged mice". *Blood* 117: pp. 2227–36. ISSN: 0006-4971. DOI: 10.1182/blood-2010-09-306019.

Chung, Y. L., Lee, M. Y., Wang, A. J., and Yao, L. F. (2003). "A therapeutic strategy uses histone deacetylase inhibitors to modulate the expression of genes involved in the pathogenesis of rheumatoid arthritis". *Mol Ther* 8: pp. 707–17. ISSN: 1525-0016 (Print) 1525-0016.

Ciccacci, Cinzia, Perricone, Carlo, Ceccarelli, Fulvia, Rufini, Sara, Di Fusco, Davide, Alessandri, Cristiano, Spinelli, Francesca Romana, Cipriano, Enrica, Novelli, Giuseppe, Valesini, Guido, et al. (2014). "A Multilocus Genetic Study in a Cohort of Italian SLE Patients Confirms the Association with STAT4 Gene and Describes a New Association with HCP5 Gene". *PLoS ONE* 9: e111991. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0111991. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4219822/.

Ciccia, F., Bombardieri, M., Principato, A., Giardina, A., Tripodo, C., Porcasi, R., Peralta, S., Franco, V., Giardina, E., Craxi, A., et al. (2009). "Overexpression of interleukin-23, but not interleukin-17, as an immunologic signature of subclinical intestinal inflammation in ankylosing spondylitis". *Arthritis Rheum* 60: pp. 955–65. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/art.24389.

Claussnitzer, M., Dankel, S. N., Kim, K. H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I. S., Beaudry, J. L., Puviindran, V., et al. (2015). "FTO Obesity Variant Circuitry and Adipocyte Browning in Humans". *N Engl J Med* 373: pp. 895–907. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1502214.

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., and Lawrence, J. B. (2009). "An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles". *Mol Cell* 33: pp. 717–26. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2009.01.026.

Coccia, E. M., Cicala, C., Charlesworth, A., Ciccarelli, C., Rossi, G. B., Philipson, L., and Sorrentino, V. (1992). "Regulation and expression of a growth arrest-specific gene (gas5) during growth, differentiation, and development". *Mol Cell Biol* 12: pp. 3514–21. ISSN: 0270-7306 (Print) 0270-7306.

Collier, S. P., Collins, P. L., Williams, C. L., Boothby, M. R., and Aune, T. M. (2012). "Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells". *J Immunol* 189: pp. 2084–8. ISSN: 0022-1767. DOI: 10.4049/jimmunol.1200774.

Colonna, Marco (2007). "All roads lead to CARD9". *Nature immunology* 8: pp. 554–555. ISSN: 1529-2908. DOI: 10.1038/ni0607-554. URL: http://dx.doi.org/10.1038/ni0607-554.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013). "Multiplex genome engineering using CRISPR/Cas systems". *Science* 339: pp. 819–23. ISSN: 0036-8075. DOI: 10.1126/science.1231143.

Conigliaro, Paola, Scrivo, Rossana, Valesini, Guido, and Perricone, Roberto (2011). "Emerging role for NK cells in the pathogenesis of inflammatory arthropathies". *Autoimmunity Reviews* 10: pp. 577–581. ISSN: 1568-9972. DOI: http://dx.doi.org/10.1016/j.autrev.2011.04.017. URL: http://www.sciencedirect.com/science/article/pii/S1568997211000851.

Consortium, The, Maller, Julian B., McVean, Gilean, Byrnes, Jake, Vukcevic, Damjan, Palin, Kimmo, Su, Zhan, Howson, Joanna M. M., Auton, Adam, Myers, Simon, et al. (2012). "Bayesian refinement of association signals for 14 loci in 3 common diseases". *Nature Genetics* 44: pp. 1294–1301. ISSN: 1061-4036. DOI: 10.1038/ng.2435. URL: http://dx.doi.org/10.1038/ng.2435.

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., et al. (2016). "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution". *Nat Genet* 48: pp. 1193–203. ISSN: 1061-4036. DOI: 10.1038/ng.3646.

Cordell, H. J., Han, Y., Mells, G. F., Li, Y., Hirschfield, G. M., Greene, C. S., Xie, G., Juran, B. D., Zhu, D., Qian, D. C., et al. (2015). "International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways". *Nat Commun* 6: p. 8019. ISSN: 2041-1723. DOI: 10.1038/ncomms9019.

Cortes, A. and Brown, M. A. (2011). "Promise and pitfalls of the Immunochip". *Arthritis Res Ther* 13: p. 101. ISSN: 1478-6354. DOI: 10.1186/ar3204.

Cortes, A., Hadler, J., Pointon, J. P., Robinson, P. C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., lee, S., et al. (2013). "Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci". *Nat Genet* 45: pp. 730–8. ISSN: 1061-4036 (Print) 1546-1718 (Electronic). DOI: 10.1038/ng.2667.

Cortes, A., Maksymowych, W. P., Wordsworth, B. P., Inman, R. D., Danoy, P., Rahman, P., Stone, M. A., Corr, M., Gensler, L. S., Gladman, D., et al. (2015a). "Association study of genes related to bone formation and resorption and the extent of radiographic change in ankylosing spondylitis". *Ann Rheum Dis* 74: pp. 1387–93. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2013-204835.

Cortes, A., Pulit, S. L., Leo, P. J., Pointon, J. J., Robinson, P. C., Weisman, M. H., Ward, M., Gensler, L. S., Zhou, X., Garchon, H. J., et al. (2015b). "Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1". *Nat Commun* 6: p. 7146. ISSN: 2041-1723. DOI: 10.1038/ncomms8146.

Cosmi, L., De Palma, R., Santarlasci, V., Maggi, L., Capone, M., Frosali, F., Rodolico, G., Querci, V., Abbate, G., Angeli, R., et al. (2008). "Human interleukin 17-producing cells originate from a CD161+CD4+ T cell precursor". *J Exp Med* 205: pp. 1903–16. ISSN: 0022-1007. DOI: 10.1084/jem.20080397.

Costantino, F., Talpin, A., Evnouchidou, I., Kadi, A., Leboime, A., Said-Nahal, R., Bonilla, N., Letourneur, F., Leturcq, T., Ka, Z., et al. (2015). "ERAP1 Gene Expression Is Influenced by Nonsynonymous Polymorphisms Associated With Predisposition to Spondyloarthritis". *Arthritis Rheumatol* 67: pp. 1525–34. ISSN: 2326-5191. DOI: 10.1002/art.39072.

Costantino, F., Chaplais, E., Leturcq, T., Said-Nahal, R., Leboime, A., Zinovieva, E., Zelenika, D., Gut, I., Charon, C., Chiocchia, G., et al. (2016). "Whole-genome single nucleotide polymorphism-based linkage analysis in spondyloarthritis multiplex families reveals a new susceptibility locus in 13q13". *Ann Rheum Dis* 75: pp. 1380–5. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2015-207720.

Costantino, F., Talpin, A., Said-Nahal, R., Leboime, A., Zinovieva, E., Zelenika, D., Gut, I., Charon, C., Izac, B., Weissman, M., et al. (2017). "A family-based genome-wide association study reveals an association of spondyloarthritis with MAPK14". *Ann Rheum Dis* 76: pp. 310–314. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2016-209449.

Costello, M. E., Ciccia, F., Willner, D., Warrington, N., Robinson, P. C., Gardiner, B., Marshall, M., Kenna, T. J., Triolo, G., and Brown, M. A. (2015). "Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis". *Arthritis Rheumatol* 67: pp. 686–691. ISSN: 2326-5191. DOI: 10.1002/art.38967.

Coulthard, S. A., Berry, P., McGarrity, S., McLaughlin, S., Ansari, A., and Redfern, C. P. F. (2017). "Azathioprine with Allopurinol: Lower Deoxythioguanosine in DNA and Transcriptome Changes Indicate Mechanistic Differences to Azathioprine Alone". *Inflamm Bowel Dis* 23: pp. 946–955. ISSN: 1078-0998. DOI: 10.1097/mib.0000000000001131.

Cowling, P., Ebringer, R., and Ebringer, A. (1980). "Association of inflammation with raised serum IgA in ankylosing spondylitis". *Ann Rheum Dis* 39: pp. 545–9. ISSN: 0003-4967 (Print) 0003-4967.

Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I., and Fakan, S. (2006). "Chromosome territories–a functional nuclear landscape". *Curr Opin Cell Biol* 18: pp. 307–16. ISSN: 0955-0674 (Print) 0955-0674. DOI: 10.1016/j.ceb.2006.04.007.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state". *Proc Natl Acad Sci U S A* 107: pp. 21931–6. ISSN: 0027-8424. DOI: 10.1073/pnas.1016071107.

Cuda, C. M., Misharin, A. V., Khare, S., Saber, R., Tsai, F., Archer, A. M., Homan, P. J., Haines G. K., 3rd, Hutcheson, J., Dorfleutner, A., et al. (2015). "Conditional deletion of caspase-8 in macrophages alters macrophage activation in a RIPK-dependent manner". *Arthritis Res Ther* 17: p. 291. ISSN: 1478-6354. DOI: 10.1186/s13075-015-0794-z.

Dabritz, J., Friedrichs, F., Weinhage, T., Hampe, J., Kucharzik, T., Lugering, A., Broeckel, U., Schreiber, S., Spieker, T., Stoll, M., et al. (2011). "The functional -374T/A polymorphism of the receptor for advanced glycation end products may modulate Crohn's disease". *Am J Physiol Gastrointest Liver Physiol* 300: G823–32. ISSN: 0193-1857. DOI: 10.1152/ajpgi.00115.2010.

Dai, Y., Rahmani, M., Dent, P., and Grant, S. (2005). "Blockade of histone deacetylase inhibitor-induced RelA/p65 acetylation and NF-kappaB activation potentiates apoptosis in leukemia cells through a process mediated by oxidative damage, XIAP downregulation, and c-Jun N-terminal kinase 1 activation". *Mol Cell Biol* 25: pp. 5429–44. ISSN: 0270-7306 (Print) 0270-7306. DOI: 10.1128/mcb.25.13.5429-5444.2005.

Dailey, Lisa (2015). "High throughput technologies for the functional discovery of mammalian enhancers: New approaches for understanding transcriptional regulatory network dynamics". *Genomics*. ISSN: 0888-7543. DOI: 10.1016/j.ygeno.2015.06.004. URL: http://dx.doi.org/10.1016/j.ygeno.2015.06.004.

Danoy, Patrick, Pryce, Karena, Hadler, Johanna, Bradbury, Linda A., Farrar, Claire, Pointon, Jennifer, Consortium, Australo-Anglo-American, Ward, Michael, Weisman, Michael, Reveille, John D., et al. (2010). "Association of variants at 1q32 and STAT3 with ankylosing spondylitis suggests genetic overlap with Crohn's disease". *PLoS genetics* 6. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1001195. URL: http://dx.doi.org/10.1371/journal.pgen.1001195.

Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., and Aerts, S. (2015). "Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling". *PLoS Genet* 11: e1004994. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1004994.

Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R. A., Milosavljevic, A., Troakes, C., et al. (2012). "Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood". *Genome Biol* 13: R43. ISSN: 1474-7596. DOI: 10.1186/gb-2012-13-6-r43.

De Rycke, L., Kruithof, E., Vandooren, B., Tak, P. P., and Baeten, D. (2006). "Pathogenesis of spondyloarthritis: insights from synovial membrane studies". *Curr Rheumatol Rep* 8: pp. 275–82. ISSN: 1523-3774 (Print) 1523-3774.

De Santa, F., Totaro, M. G., Prosperini, E., Notarbartolo, S., Testa, G., and Natoli, G. (2007). "The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing". *Cell* 130: pp. 1083–94. ISSN: 0092-8674 (Print) 0092-8674. DOI: 10.1016/j.cell.2007.08.019.

Dean, Erin (2016). "Ankylosing spondylitis". *Nursing standard (Royal College of Nursing (Great Britain) : 1987)* 31: p. 15. ISSN: 0029-6570. DOI: 10.7748/ns.31.7.15.s16. URL: http://dx.doi.org/10.7748/ns.31.7.15.s16.

Degner, Jacob F., Pai, Athma A., Pique-Regi, Roger, Veyrieras, Jean-Baptiste B., Gaffney, Daniel J., Pickrell, Joseph K., De Leon, Sherryl, Michelini, Katelyn, Lewellen, Noah, Crawford, Gregory E., et al. (2012). "DNaseI sensitivity QTLs are a major determinant of human expression variation". *Nature* 482: pp. 390–394. ISSN: 0028-0836. DOI: 10.1038/nature10808. URL: http://dx.doi.org/10.1038/nature10808.

Dehghan, A., Dupuis, J., Barbalic, M., Bis, J. C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P., et al. (2011). "Meta-analysis of genome-wide association studies in ¿80 000 subjects identifies multiple loci for C-reactive protein levels". *Circulation* 123: pp. 731–8. ISSN: 0009-7322. DOI: 10.1161/circulationaha.110.948570.

Delaneau, O., Marchini, J., and Zagury, J. F. (2011). "A linear complexity phasing method for thousands of genomes". *Nat Methods* 9: pp. 179–81. ISSN: 1548-7091. DOI: 10.1038/nmeth.1785.

Deplancke, B., Alpern, D., and Gardeux, V. (2016). "The Genetics of Transcription Factor DNA Binding Variation". *Cell* 166: pp. 538–54. ISSN: 0092-8674. DOI: 10.1016/j.cell.2016.07.012.

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A. Y., Dixon, J., Maliskova, L., Guan, K. L., Shen, Y., and Ren, B. (2016). "A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening". *Genome Res* 26: pp. 397–405. ISSN: 1088-9051. DOI: 10.1101/gr.197152.115.

Dimmer, Emily C., Huntley, Rachael P., Barrell, Daniel G., Binns, David, Draghici, Sorin, Camon, Evelyn B., Hubank, Mike, Talmud, Philippa J., Apweiler, Rolf, and Lovering, Ruth C. (2008). "The Gene Ontology Providing a Functional Role in Proteomic Studies". *PROTEOMICS* 8: n/a–n/a. ISSN: 1615-9861. DOI: 10.1002/pmic.200800002. URL: http://dx.doi.org/10.1002/pmic.200800002.

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). "Chromatin architecture reorganization during stem cell differentiation". *Nature* 518: pp. 331–6. ISSN: 0028-0836. DOI: 10.1038/nature14222.

Doench, J. G. and Sharp, P. A. (2004). "Specificity of microRNA target selection in translational repression". *Genes Dev* 18: pp. 504–11. ISSN: 0890-9369 (Print) 0890-9369. DOI: 10.1101/gad.1184404.

Dolcino, M., Ottria, A., Barbieri, A., Patuzzo, G., Tinazzi, E., Argentino, G., Beri, R., Lunardi, C., and Puccetti, A. (2015). "Gene Expression Profiling in Peripheral Blood Cells and Synovial Membranes of Patients with Psoriatic Arthritis". *PLoS One* 10: e0128262. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0128262.

Dolcino, M., Tinazzi, E., Pelosi, A., Patuzzo, G., Moretta, F., Lunardi, C., and Puccetti, A. (2017). "Gene Expression Analysis before and after Treatment with Adalimumab in Patients with Ankylosing Spondylitis Identifies Molecular Pathways Associated with Response to Therapy". *Genes (Basel)* 8. ISSN: 2073-4425 (Print) 2073-4425. DOI: 10.3390/genes8040127.

Dranoff, G. (2004). "Cytokines in cancer pathogenesis and cancer therapy". *Nat Rev Cancer* 4: pp. 11–22. ISSN: 1474-175X (Print) 1474-175x. DOI: 10.1038/nrc1252.

Duan, R., Leo, P., Bradbury, L., Brown, M. A., and Thomas, G. (2010). "Gene expression profiling reveals a downregulation in immune-associated genes in patients with AS". *Ann Rheum Dis* 69: pp. 1724–9. ISSN: 0003-4967. DOI: 10.1136/ard.2009.111690.

Dunham Ian, The ENCODE Consortium (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489: pp. 57–74. ISSN: 0028-0836. DOI: 10.1038/nature11247.

Ecker, S., Chen, L., Pancaldi, V., Bagger, F. O., Fernandez, J. M., Santa Pau, E. Carrillo de, Juan, D., Mann, A. L., Watt, S., Casale, F. P., et al. (2017). "Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types". *Genome Biol* 18: p. 18. ISSN: 1474-7596. DOI: 10.1186/s13059-017-1156-8.

Eckmann, L., Nebelsiek, T., Fingerle, A. A., Dann, S. M., Mages, J., Lang, R., Robine, S., Kagnoff, M. F., Schmid, R. M., Karin, M., et al. (2008). "Opposing functions of IKKbeta during acute and chronic intestinal inflammation". *Proc Natl Acad Sci U S A* 105: pp. 15058–63. ISSN: 0027-8424. DOI: 10.1073/pnas.0808216105.

Eder, L., Chandran, V., Shen, H., Cook, R. J., and Gladman, D. D. (2010). "Is ASDAS better than BASDAI as a measure of disease activity in axial psoriatic arthritis?" *Ann Rheum Dis* 69: pp. 2160–4. ISSN: 0003-4967. DOI: 10.1136/ard.2010.129726.

Edwards, Stacey L., Beesley, Jonathan, French, Juliet D., and Dunning, Alison M. (2013). "Beyond GWASs: illuminating the dark road from association to function". *American journal of human genetics* 93: pp. 779–797. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2013.10.012. URL: http://dx.doi.org/10.1016/j.ajhg.2013.10.012.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., and Nadeau, J. H. (2010). "Missing heritability and strategies for finding the underlying causes of complex disease". *Nat Rev Genet* 11: pp. 446–50. ISSN: 1471-0056. DOI: 10.1038/nrg2809.

Elling, R., Chan, J., and Fitzgerald, K. A. (2016). "Emerging role of long noncoding RNAs as regulators of innate immune cell development and inflammatory gene expression". *Eur J Immunol* 46: pp. 504–12. ISSN: 0014-2980. DOI: 10.1002/eji.201444558.

Ellinghaus, D., Jostins, L., Spain, S. L., and Cortes, A. (2016). "Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci". *Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci*. URL: http://scholar.google.com/scholar?q=Analysisoffivechronicinflammatorydiseasesidentifies27newassociationsandhighspecificpatternsatsharedloci&btnG=&hl=en&num=20&as_sdt=0%2C22.

Ernst, Jason and Kellis, Manolis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome". *Nature biotechnology* 28: pp. 817–825. ISSN: 1087-0156. DOI: 10.1038/nbt.1662. URL: http://dx.doi.org/10.1038/nbt.1662.

Ernst, Jason and Kellis, Manolis (2012). "ChromHMM: automating chromatin-state discovery and characterization". *Nature Methods* 9: pp. 215–216. ISSN: 1548-7091. DOI: 10.1038/nmeth.1906. URL: http://dx.doi.org/10.1038/nmeth.1906.

Ernst, Jason, Kheradpour, Pouya, Mikkelsen, Tarjei S., Shoresh, Noam, Ward, Lucas D., Epstein, Charles B., Zhang, Xiaolan, Wang, Li, Issner, Robbyn, Coyne, Michael, et al. (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types". *Nature* 473: pp. 43–49. ISSN: 0028-0836. DOI: 10.1038/nature09906. URL: http://dx.doi.org/10.1038/nature09906.

Evans, D. M., Spencer, C. C., Pointon, J. J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A., Pirinen, M., Stone, M. A., et al. (2011). "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility". *Nat Genet* 43: pp. 761–7. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.873.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., et al. (2016). "The Reactome pathway Knowledgebase". *Nucleic Acids Res* 44: pp. D481–7. ISSN: 0305-1048. DOI: 10.1093/nar/gkv1351.

Fairfax, B. P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F. O., and Knight, J. C. (2012). "Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles". *Nat Genet* 44: pp. 502–10. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.2205.

Fairfax, Benjamin P. and Knight, Julian C. (2014). "Genetics of gene expression in immunity to infection". *Current opinion in immunology* 30: pp. 63–71. ISSN: 0952-7915. DOI: 10.1016/j.coi.2014.07.001. URL: http://dx.doi.org/10.1016/j.coi.2014.07.001.

Fairfax, Benjamin P., Humburg, Peter, Makino, Seiko, Naranbhai, Vivek, Wong, Daniel, Lau, Evelyn, Jostins, Luke, Plant, Katharine, Andrews, Robert, McGee, Chris, et al. (2014). "Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression". *Science* 343. DOI: 10.1126/science.1246949. URL: http://www.sciencemag.org/content/343/6175/1246949.abstract.

Fang, F., Pan, J., Xu, L., Li, G., and Wang, J. (2015). "Identification of Potential Transcriptomic Markers in Developing Ankylosing Spondylitis: A Meta-Analysis of Gene Expression Profiles". *Identification of Potential Transcriptomic Markers in Developing Ankylosing Spondylitis: A Meta-Analysis of Gene Expression Profiles*. ISSN: 2314-6133. DOI: 10.1155/2015/826316. URL: http://dx.doi.org/10.1155/2015/826316.

Fang, Hai, Knezevic, Bogdan, Burnham, Katie L., and Knight, Julian C. (2016). "XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits". *Genome medicine* 8: p. 129. ISSN: 1756-994X. DOI: 10.1186/s13073-016-0384-y. URL: http://dx.doi.org/10.1186/s13073-016-0384-y.

Farh, Kyle K., Marson, Alexander, Zhu, Jiang, Kleinewietfeld, Markus, Housley, William J., Beik, Samantha, Shoresh, Noam, Whitton, Holly, Ryan, Russell J., Shishkin, Alexander A., et al. (2015). "Genetic and epigenetic fine mapping of causal autoimmune disease variants". *Nature* 518: pp. 337–343. ISSN: 0028-0836. DOI: 10.1038/nature13835. URL: http://dx.doi.org/10.1038/nature13835.

Feinberg, A. P. (2007). "Phenotypic plasticity and the epigenetics of human disease". *Nature* 447: pp. 433–40. ISSN: 0028-0836. DOI: 10.1038/nature05919.

Feltham, Rebecca, Vince, James E., and Lawlor, Kate E. (2017). "Caspase-8: not so silently deadly". *Clinical & Translational Immunology* 6: e124. ISSN: 2050-0068. DOI: 10.1038/cti.2016.83. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5292560/.

Fert, I., Cagnard, N., Glatigny, S., Letourneur, F., Jacques, S., Smith, J. A., Colbert, R. A., Taurog, J. D., Chiocchia, G., Araujo, L. M., et al. (2014). "Reverse interferon signature is characteristic of antigen-presenting cells in human and rat spondyloarthritis". *Arthritis Rheumatol* 66: pp. 841–51. ISSN: 2326-5191. DOI: 10.1002/art.38318.

Fischer, R., Trudgian, D. C., Wright, C., Thomas, G., Bradbury, L. A., Brown, M. A., Bowness, P., and Kessler, B. M. (2012). "Discovery of candidate serum proteomic and metabolomic biomarkers in ankylosing spondylitis". *Mol Cell Proteomics* 11: p. M111.013904. ISSN: 1535-9476. DOI: 10.1074/mcp.M111.013904.

Fitzgerald, K. A., McWhirter, S. M., Faia, K. L., Rowe, D. C., Latz, E., Golenbock, D. T., Coyle, A. J., Liao, S. M., and Maniatis, T. (2003). "IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway". *Nat Immunol* 4: pp. 491–6. ISSN: 1529-2908 (Print) 1529-2908. DOI: 10.1038/ni921.

Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., and Ahmad, T. (2010). "Genome-wide meta-analysis increases to 71 the number of confirmed Crohns disease susceptibility loci". *Nat Genet* 42. DOI: 10.1038/ng.717. URL: https://doi.org/10.1038/ng.717.

Fraser, James, Williamson, Iain, Bickmore, Wendy A., and Dostie, Jose (2015). "An Overview of Genome Organization and How We Got There: from FISH to Hi-C". *Microbiology and molecular biology reviews : MMBR* 79: pp. 347–372. ISSN: 1092-2172. DOI: 10.1128/MMBR.00006-15. URL: http://dx.doi.org/10.1128/MMBR.00006-15.

Gaffney, D. J. (2013). "Global properties and functional complexity of human gene regulatory variation". *PLoS Genet* 9: e1003501. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1003501.

Garcia, F., Rognan, D., Lamas, J. R., Marina, A., and Castro, J. A. Lopez de (1998). "An HLA-B27 polymorphism (B*2710) that is critical for T-cell recognition has limited effects on peptide specificity". *Tissue Antigens* 51: pp. 1–9. ISSN: 0001-2815 (Print) 0001-2815.

Garcia-Fernandez, S., Gonzalez, S., Mina Blanco, A., Martinez-Borra, J., Blanco-Gelaz, M., Lopez-Vazquez, A., and Lopez-Larrea, C. (2001). "New insights regarding HLA-B27 diversity in the Asian population". *Tissue Antigens* 58: pp. 259–62. ISSN: 0001-2815 (Print) 0001-2815.

Gassen, K. L. van, Wit, M. de, Koerkamp, M. J., Rensen, M. G., Rijen, P. C. van, Holstege, F. C., Lindhout, D., and Graan, P. N. de (2008). "Possible role of the innate immunity in temporal lobe epilepsy". *Epilepsia* 49: pp. 1055–65. ISSN: 0013-9580. DOI: 10.1111/j.1528-1167.2007.01470.x.

Gaulton, Kyle J., Ferreira, Teresa, Lee, Yeji, Raimondo, Anne, Mgi, Reedik, Reschen, Michael E., Mahajan, Anubha, Locke, Adam, William Rayner, N., Robertson, Neil, et al. (2015). "Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci". *Nature genetics*. ISSN: 1061-4036. DOI: 10.1038/ng.3437. URL: http://dx.doi.org/10.1038/ng.3437.

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007). "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin". *Genome Res* 17. DOI: 10.1101/gr.5533506. URL: http://dx.doi.org/10.1101/gr.5533506.

Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). "Finding genes that underlie complex traits". *Science* 298: pp. 2345–9. ISSN: 0036-8075. DOI: 10.1126/science.1076641.

Gomez, J. A., Wapinski, O. L., Yang, Y. W., Bureau, J. F., Gopinath, S., Monack, D. M., Chang, H. Y., Brahic, M., and Kirkegaard, K. (2013). "The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus". *Cell* 152: pp. 743–54. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.01.015.

Gonzalez, S., Martinez-Borra, J., and Lopez-Larrea, C. (1999). "Immunogenetics, HLA-B27 and spondyloarthropathies". *Curr Opin Rheumatol* 11: pp. 257–64. ISSN: 1040-8711 (Print) 1040-8711 (Linking).

Govindan, G. and Ramalingam, S. (2016). "Programmable Site-Specific Nucleases for Targeted Genome Engineering in Higher Eukaryotes". *J Cell Physiol* 231: pp. 2380–92. ISSN: 0021-9541. DOI: 10.1002/jcp.25367.

Gregory, Adam P., Dendrou, Calliope A., Attfield, Kathrine E., Haghikia, Aiden, Xifara, Dionysia K., Butter, Falk, Poschmann, Gereon, Kaur, Gurman, Lambert, Lydia, Leach, Oliver A., et al. (2012). "TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis". *Nature* 488: pp. 508–511. ISSN: 0028-0836. DOI: 10.1038/nature11307. URL: http://dx.doi.org/10.1038/nature11307.

Greten, F. R., Arkan, M. C., Bollrath, J., Hsu, L. C., Goode, J., Miething, C., Goktuna, S. I., Neuenhahn, M., Fierer, J., Paxian, S., et al. (2007). "NF-kappaB is a negative regulator of IL-1beta secretion as revealed by genetic and pharmacological inhibition of IKKbeta". *Cell* 130: pp. 918–31. ISSN: 0092-8674 (Print) 0092-8674. DOI: 10.1016/j.cell.2007.07.009.

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing". *Mol Cell* 27: pp. 91–105. ISSN: 1097-2765 (Print) 1097-2765. DOI: 10.1016/j.molcel.2007.06.017.

Gu, J., Marker-Hermann, E., Baeten, D., Tsai, W. C., Gladman, D., Xiong, M., Deister, H., Kuipers, J. G., Huang, F., Song, Y. W., et al. (2002). "A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients". *Rheumatology (Oxford)* 41: pp. 759–66. ISSN: 1462-0324 (Print) 1462-0324.

Gu, Jieruo, Wei, Yu, Wei, James C. C., Huang, Feng, Jan, Ming, Centola, Michael, Frank, Mark B., and Yu, David (2009). "Identification of RGS1 as a candidate biomarker for undifferentiated spondylarthritis by genome wide expression profiling and real time polymerase chain reaction". *Arthritis & Rheumatism* 60: pp. 3269–3279. DOI: 10.1002/art.24968. URL: http://dx.doi.org/10.1002/art.24968.

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels". *Nature* 466: pp. 835–40. ISSN: 0028-0836. DOI: 10.1038/nature09267.

Guo, Ya, Xu, Quan, Canzio, Daniele, Shou, Jia, Li, Jinhuan, Gorkin, David U., Jung, Inkyung, Wu, Haiyang, Zhai, Yanan, Tang, Yuanxiao, et al. (2015). "CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function". *Cell* 162: pp. 900–910. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.07.038. URL: http://dx.doi.org/10.1016/j.cell.2015.07.038.

Gusev, Alexander, Lee, Hong S., Trynka, Gosia, Finucane, Hilary, Vilhjlmsson, Bjarni J., Xu, Han, Zang, Chongzhi, Ripke, Stephan, Bulik-Sullivan, Brendan, Stahl, Eli, et al. (2014). "Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases". *American journal of human genetics* 95: pp. 535–552. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2014.10.004. URL: http://dx.doi.org/10.1016/j.ajhg.2014.10.004.

Haberman, Y., Tickle, T. L., Dexheimer, P. J., Kim, M. O., Tang, D., Karns, R., Baldassano, R. N., Noe, J. D., Rosh, J., Markowitz, J., et al. (2014). "Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature". *J Clin Invest* 124: pp. 3617–33. ISSN: 0021-9738. DOI: 10.1172/jci75436.

Hagege, Helene, Klous, Petra, Braem, Caroline, Splinter, Erik, Dekker, Job, Cathala, Guy, Laat, Wouter de, and Forne, Thierry (2007). "Quantitative analysis of chromosome conformation capture assays (3C-qPCR)". *Nat. Protocols* 2: pp. 1722–1733. ISSN: 1754-2189. URL: http://dx.doi.org/10.1038/nprot.2007.243.

Hal, N. L. van, Vorst, O., Houwelingen, A. M. van, Kok, E. J., Peijnenburg, A., Aharoni, A., Tunen, A. J. van, and Keijer, J. (2000). "The application of DNA microarrays in gene expression analysis". *J Biotechnol* 78: pp. 271–80. ISSN: 0168-1656 (Print) 0168-1656.

Hammer, G. E., Gonzalez, F., Champsaur, M., Cado, D., and Shastri, N. (2006). "The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules". *Nat Immunol* 7: pp. 103–12. ISSN: 1529-2908 (Print) 1529-2908. DOI: 10.1038/ni1286.

Hammer, G. E., Turer, E. E., Taylor, K. E., Fang, C. J., Advincula, R., Oshima, S., Barrera, J., Huang, E. J., Hou, B., Malynn, B. A., et al. (2011). "Expression of A20 by dendritic cells preserves immune homeostasis and prevents colitis and spondyloarthritis". *Nat Immunol* 12: pp. 1184–93. ISSN: 1529-2908. DOI: 10.1038/ni.2135.

Hansen, T., Olsen, L., Lindow, M., Jakobsen, K. D., Ullum, H., Jonsson, E., Andreassen, O. A., Djurovic, S., Melle, I., Agartz, I., et al. (2007). "Brain expressed microRNAs implicated in schizophrenia etiology". *PLoS One* 2: e873. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000873.

Haraksingh, R. R. and Snyder, M. P. (2013). "Impacts of variation in the human genome on gene regulation". *J Mol Biol* 425: pp. 3970–7. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2013.07.015.

Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., Ren, B., Fu, X. D., Topol, E. J., Rosenfeld, M. G., et al. (2011). "9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response". *Nature* 470: pp. 264–8. ISSN: 0028-0836. DOI: 10.1038/nature09753.

Harvey, D., Pointon, J. J., Evans, D. M., Karaderi, T., Farrar, C., Appleton, L. H., Sturrock, R. D., Stone, M. A., Oppermann, U., Brown, M. A., et al. (2009). "Investigating the genetic association between ERAP1 and ankylosing spondylitis". *Hum Mol Genet* 18: pp. 4204–12. ISSN: 1460-2083 (Electronic) 0964-6906 (Linking). DOI: 10.1093/hmg/ddp371.

Haynes K. T. Glazov, E. Matthew A. Brown and Thomas, Gethin (2014). "[Abstract]A Novel Monocyte-Specific Transcript Underlies the Chromosome 21q22 Intergenic Genetic Association in Ankylosing Spondylitis". *UNPUBLISHED*.

Haywood, M. E., Rose, S. J., Horswell, S., Lees, M. J., Fu, G., Walport, M. J., and Morley, B. J. (2006). "Overlapping BXSB congenic intervals, in combination with microarray gene expression, reveal novel lupus candidate genes". *Genes Immun* 7: pp. 250–63. ISSN: 1466-4879 (Print) 1466-4879. DOI: 10.1038/sj.gene.6364294.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". *Nat Genet* 39: pp. 311–8. ISSN: 1061-4036 (Print) 1061-4036. DOI: 10.1038/ng1966.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., and Harp, L. F. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression". *Nature* 459. DOI: 10.1038/nature07829. URL: https://doi.org/10.1038/nature07829.

Heinz, Sven, Romanoski, Casey E., Benner, Christopher, and Glass, Christopher K. (2015). "The selection and function of cell type-specific enhancers". *Nature reviews. Molecular cell biology*. ISSN: 1471-0072. DOI: 10.1038/nrm3949. URL: http://dx.doi.org/10.1038/nrm3949.

Heyn, H. (2014). "A symbiotic liaison between the genetic and epigenetic code". *Front Genet* 5: p. 113. ISSN: 1664-8021 (Print) 1664-8021. DOI: 10.3389/fgene.2014.00113.

Heyn, H., Sayols, S., Moutinho, C., Vidal, E., Sanchez-Mut, J. V., Stefansson, O. A., Nadal, E., Moran, S., Eyfjord, J. E., Gonzalez-Suarez, E., et al. (2014). "Linkage of DNA methylation quantitative trait loci to human cancer risk". *Cell Rep* 7: pp. 331–8. DOI: 10.1016/j.celrep.2014.03.016.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., and Collins, F. S. (2009). "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits". *Proc Natl Acad Sci U S A* 106. DOI: 10.1073/pnas.0903103106. URL: https://doi.org/10.1073/pnas.0903103106.

Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Benard, M., Fox, A. H., et al. (2014). "NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies". *Mol Biol Cell* 25: pp. 169–83. ISSN: 1059-1524. DOI: 10.1091/mbc.E13-09-0558.

Hmamouchi, I., Bahiri, R., and Hajjaj-Hassouni, N. (2011). "Clinical and radiological presentations of late-onset spondyloarthritis". *ISRN Rheumatol* 2011: p. 840475. ISSN: 2090-5467. DOI: 10.5402/2011/840475.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., Hoke, H. A., and Young, R. A. (2013). "Super-enhancers in the control of cell identity and disease". *Cell* 155: pp. 934–47. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.09.053.

Hofmann, M. A., Drury, S., Hudson, B. I., Gleason, M. R., Qu, W., Lu, Y., Lalla, E., Chitnis, S., Monteiro, J., Stickland, M. H., et al. (2002). "RAGE and arthritis: the G82S polymorphism amplifies the inflammatory response". *Genes Immun* 3: pp. 123–35. ISSN: 1466-4879 (Print) 1466-4879. DOI: 10.1038/sj.gene.6363861.

Hong, S. N., Park, C., Park, S. J., Lee, C. K., Ye, B. D., Kim, Y. S., Lee, S., Chae, J., Kim, J. I., and Kim, Y. H. (2016). "Deep resequencing of 131 Crohn's disease associated genes in pooled DNA confirmed three reported variants and identified eight novel variants". *Gut* 65: pp. 788–96. ISSN: 0017-5749. DOI: 10.1136/gutjnl-2014-308617.

Horiuchi, T., Mitoma, H., Harashima, S., Tsukamoto, H., and Shimoda, T. (2010). "Transmembrane TNF-alpha: structure, function and interaction with anti-TNF agents". *Rheumatology (Oxford)* 49: pp. 1215–28. ISSN: 1462-0324. DOI: 10.1093/rheumatology/keq031.

Hovelmeyer, N., Reissig, S., Xuan, N. T., Adams-Quack, P., Lukas, D., Nikolaev, A., Schluter, D., and Waisman, A. (2011). "A20 deficiency in B cells enhances B-cell proliferation and results in the development of autoantibodies". *Eur J Immunol* 41: pp. 595–601. ISSN: 0014-2980. DOI: 10.1002/eji.201041313.

Howie, B. N., Donnelly, P., and Marchini, J. (2009). "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies". *PLoS Genet* 5. DOI: 10.1371/journal.pgen.1000529. URL: https://doi.org/10.1371/journal.pgen.1000529.

Hrdlickova, Barbara, Kumar, Vinod, Kanduri, Kartiek, Zhernakova, Daria V., Tripathi, Subhash, Karjalainen, Juha, Lund, Riikka J., Li, Yang, Ullah, Ubaid, Modderman, Rutger, et al. (2014). "Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity". *Genome Medicine* 6: p. 88. ISSN: 1756-994X. DOI: 10.1186/s13073-014-0088-0. URL: http://dx.doi.org/10.1186/s13073-014-0088-0.

Huang, J., Song, G., Yin, Z., Luo, X., and Ye, Z. (2014). "Elevated miR-29a expression is not correlated with disease activity index in PBMCs of patients with ankylosing spondylitis". *Mod Rheumatol* 24: pp. 331–4. ISSN: 1439-7595. DOI: 10.3109/14397595.2013.854077.

Hughes, Jim R., Roberts, Nigel, McGowan, Simon, Hay, Deborah, Giannoulatou, Eleni, Lynch, Magnus, De Gobbi, Marco, Taylor, Stephen, Gibbons, Richard, and Higgs, Douglas R. (2014). "Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment". *Nat Genet* 46: pp. 205–212. ISSN: 1061-4036. DOI: 10.1038/ng.2871http://www.nature.com/ng/journal/v46/n2/abs/ng.2871.html#supplementary-information. URL: http://dx.doi.org/10.1038/ng.2871.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., et al. (2001). "Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease". *Nature* 411: pp. 599–603. ISSN: 0028-0836 (Print) 0028-0836. DOI: 10.1038/35079107.

Hulur, Imge, Gamazon, Eric R., Skol, Andrew D., Xicola, Rosa M., Llor, Xavier, Onel, Kenan, Ellis, Nathan A., and Kupfer, Sonia S. (2015). "Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci". *BMC Genomics* 16: p. 138. ISSN: 1471-2164. DOI: 10.1186/s12864-015-1292-z. URL: https://doi.org/10.1186/s12864-015-1292-z.

Hunt, K. A., Zhernakova, A., Turner, G., Heap, G. A., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L. C., Ryan, A. W., Panesar, D., et al. (2008). "Newly identified genetic risk variants for celiac disease related to the immune response". *Nat Genet* 40: pp. 395–402. ISSN: 1061-4036. DOI: 10.1038/ng.102.

Iborra, M., Bernuzzi, F., Correale, C., Vetrano, S., Fiorino, G., Beltran, B., Marabita, F., Locati, M., Spinelli, A., Nos, P., et al. (2013). "Identification of serum and tissue micro-RNA expression profiles in different stages of inflammatory bowel disease". *Clin Exp Immunol* 173: pp. 250–8. ISSN: 0009-9104. DOI: 10.1111/cei.12104.

IMSGC (2010). "IL12A, MPHOSPH9/CDK2AP1 and RGS1 are novel multiple sclerosis susceptibility loci". *Genes Immun* 11: pp. 397–405. ISSN: 1466-4879. DOI: 10.1038/gene.2010.28.

Inman, R. D. (2009). "Innate immunity of spondyloarthritis: the role of toll-like receptors". *Adv Exp Med Biol* 649: pp. 300–9. ISSN: 0065-2598 (Print) 0065-2598.

Inman, R. D. and El-Gabalawy, H. S. (2009). "The immunology of ankylosing spondylitis and rheumatoid arthritis: a tale of similarities and dissimilarities". *Clin Exp Rheumatol* 27: S26–32. ISSN: 0392-856X (Print) 0392-856x.

Inoue, F. and Ahituv, N. (2015). "Decoding enhancers using massively parallel reporter assays". *Genomics* 106: pp. 159–64. ISSN: 0888-7543. DOI: 10.1016/j.ygeno.2015.06.005.

Ipseiz, N., Uderhardt, S., Scholtysek, C., Steffen, M., Schabbauer, G., Bozec, A., Schett, G., and Kronke, G. (2014). "The nuclear receptor Nr4a1 mediates anti-inflammatory effects of apoptotic cells". *J Immunol* 192: pp. 4852–8. ISSN: 0022-1767. DOI: 10.4049/jimmunol.1303377.

Ishikawa, S., Mima, T., Aoki, C., Yoshio-Hoshino, N., Adachi, Y., Imagawa, T., Mori, M., Tomiita, M., Iwata, N., Murata, T., et al. (2009). "Abnormal expression of the genes involved in cytokine networks and mitochondrial function in systemic juvenile idiopathic arthritis identified by DNA microarray analysis". *Ann Rheum Dis* 68: pp. 264–72. ISSN: 0003-4967. DOI: 10.1136/ard.2007.079533.

Jaakkola, E., Herzberg, I., Laiho, K., Barnardo, M. C., Pointon, J. J., Kauppi, M., Kaarela, K., Tuomilehto-Wolf, E., Tuomilehto, J., Wordsworth, B. P., et al. (2006). "Finnish HLA studies confirm the increased risk conferred by HLA-B27 homozygosity in ankylosing spondylitis". *Ann Rheum Dis* 65: pp. 775–80. ISSN: 0003-4967 (Print) 0003-4967. DOI: 10.1136/ard.2005.041103.

Jacob, A., Cooney, D., Pradhan, M., and Coggeshall, K. M. (2002). "Convergence of signaling pathways on the activation of ERK in B cells". *J Biol Chem* 277: pp. 23420–6. ISSN: 0021-9258 (Print) 0021-9258. DOI: 10.1074/jbc.M202485200.

Javierre, Biola M., Burren, Oliver S., Wilder, Steven P., Kreuzhuber, Roman, Hill, Steven M., Sewitz, Sven, Cairns, Jonathan, Wingett, Steven W., Vrnai, Csilla, Thiecke, Michiel J., et al. (2016). "Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters". *Cell* 167: p. 1369. ISSN: 0092-8674. DOI: 10.1016/j.cell.2016.09.037. URL: http://dx.doi.org/10.1016/j.cell.2016.09.037.

Jellusova, J. and Rickert, R. C. (2016). "The PI3K pathway in B cell metabolism". *Crit Rev Biochem Mol Biol* 51: pp. 359–378. ISSN: 1040-9238. DOI: 10.1080/10409238.2016.1215288.

Ji, J. D., Cha, E. S., and Lee, W. J. (2014). "Association of miR-146a polymorphisms with systemic lupus erythematosus: a meta-analysis". *Lupus* 23: pp. 1023–30. ISSN: 0961-2033. DOI: 10.1177/0961203314534512.

Jin, J., Chang, Y., and Wei, W. (2010). "Clinical application and evaluation of anti-TNF-alpha agents for the treatment of rheumatoid arthritis". *Acta Pharmacol Sin* 31: pp. 1133–40. ISSN: 1671-4083. DOI: 10.1038/aps.2010.134.

Johnson, B. A., Wang, J., Taylor, E. M., Caillier, S. J., Herbert, J., Khan, O. A., Cross, A. H., De Jager, P. L., Gourraud, P. A., Cree, B. C., et al. (2010). "Multiple sclerosis susceptibility alleles in African Americans". *Genes Immun* 11: pp. 343–50. ISSN: 1466-4879. DOI: 10.1038/gene.2009.81.

Jones, R. G. and Pearce, E. J. (2017). "MenTORing Immunity: mTOR Signaling in the Development and Function of Tissue-Resident Immune Cells". *Immunity* 46: pp. 730–742. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2017.04.028.

Jostins, Luke, Ripke, Stephan, Weersma, Rinse K., Duerr, Richard H., McGovern, Dermot P., Hui, Ken Y., Lee, James C., Philip Schumm, L., Sharma, Yashoda, Anderson, Carl A., et al. (2012). "Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease". *Nature* 491: pp. 119–124. ISSN: 0028-0836. DOI: http://www.nature.com/nature/journal/v491/n7422/abs/nature11582.html#supplementary-information. URL: http://dx.doi.org/10.1038/nature11582.

Julia, A., Domenech, E., Ricart, E., Tortosa, R., Garcia-Sanchez, V., and Gisbert, J. P. (2013). "A genome-wide association study on a southern European population identifies a new Crohns disease susceptibility locus at RBX1-EP300". *Gut* 62. DOI: 10.1136/gutjnl-2012-302865. URL: https://doi.org/10.1136/gutjnl-2012-302865.

Jung, S. H., Yim, S. H., Hu, H. J., Lee, K. H., Lee, J. H., Sheen, D. H., Lim, M. K., Kim, S. Y., Park, S. W., Kim, S. H., et al. (2014). "Genome-wide copy number variation analysis identifies deletion variants associated with ankylosing spondylitis". *Arthritis Rheumatol* 66: pp. 2103–12. ISSN: 2326-5191. DOI: 10.1002/art.38650.

Kadi, A., Costantino, F., Izac, B., Leboime, A., Said-Nahal, R., Garchon, H. J., Chiocchia, G., and Breban, M. (2013). "Brief report: the IL23R nonsynonymous polymorphism rs11209026 is associated with radiographic sacroiliitis in spondyloarthritis". *Arthritis Rheum* 65: pp. 2655–60. ISSN: 0004-3591. DOI: 10.1002/art.38060.

Kalla, R., Ventham, N. T., Kennedy, N. A., Quintana, J. F., Nimmo, E. R., Buck, A. H., and Satsangi, J. (2014). "MicroRNAs: new players in IBD". *Gut*. DOI: 10.1136/gutjnl-2014-307891. URL: http://dx.doi.org/10.1136/gutjnl-2014-307891.

Kanehisa, M. and Goto, S. (2000). "KEGG: kyoto encyclopedia of genes and genomes". *Nucleic Acids Res* 28: pp. 27–30. ISSN: 0305-1048 (Print) 0305-1048.

Kankaanranta, H., Moilanen, E., and Zhang, X. (2005). "Pharmacological regulation of human eosinophil apoptosis". *Curr Drug Targets Inflamm Allergy* 4: pp. 433–45. ISSN: 1568-010X (Print) 1568-010x.

Kankaanranta, H., Janka-Junttila, M., Ilmarinen-Salo, P., Ito, K., Jalonen, U., Ito, M., Adcock, I. M., Moilanen, E., and Zhang, X. (2010). "Histone deacetylase inhibitors induce apoptosis in human eosinophils and neutrophils". *J Inflamm (Lond)* 7: p. 9. ISSN: 1476-9255. DOI: 10.1186/1476-9255-7-9.

Karczewski, K. J., Dudley, J. T., Kukurba, K. R., Chen, R., Butte, A. J., Montgomery, S. B., and Snyder, M. (2013). "Systematic functional regulatory assessment of disease-associated variants". *Proc Natl Acad Sci U S A* 110: pp. 9607–12. ISSN: 0027-8424. DOI: 10.1073/pnas.1219099110.

Khan, M. A., Mathieu, A., Sorrentino, R., and Akkoc, N. (2007). "The pathogenetic role of HLA-B27 and its subtypes". *Autoimmun Rev* 6: pp. 183–9. ISSN: 1568-9972 (Print) 1568-9972 (Linking). DOI: 10.1016/j.autrev.2006.11.003.

Kim, Daehwan, Pertea, Geo, Trapnell, Cole, Pimentel, Harold, Kelley, Ryan, and Salzberg, Steven L. (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions". *Genome Biology* 14: R36. ISSN: 1474-760X. DOI: 10.1186/gb-2013-14-4-r36. URL: https://doi.org/10.1186/gb-2013-14-4-r36http://dx.doi.org/10.1186/gb-2013-14-4-r36.

Kim, S. W., Ramasamy, K., Bouamar, H., Lin, A. P., Jiang, D., and Aguiar, R. C. (2012). "MicroRNAs miR-125a and miR-125b constitutively activate the NF-kappaB pathway by targeting the tumor necrosis factor alpha-induced protein 3 (TNFAIP3, A20)". *Proc Natl Acad Sci U S A* 109: pp. 7865–70. ISSN: 0027-8424. DOI: 10.1073/pnas.1200081109.

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). "Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor". *Sci Signal* 3: ra8. ISSN: 1945-0877. DOI: 10.1126/scisignal.2000568.

Kleinschek, M. A., Boniface, K., Sadekova, S., Grein, J., Murphy, E. E., Turner, S. P., Raskin, L., Desai, B., Faubion, W. A., Waal Malefyt, R. de, et al. (2009). "Circulating and gut-resident human Th17 cells express CD161 and promote intestinal inflammation". *J Exp Med* 206: pp. 525–34. ISSN: 0022-1007. DOI: 10.1084/jem.20081712.

Kolodziej, L. E., Lodolce, J. P., Chang, J. E., Schneider, J. R., Grimm, W. A., Bartulis, S. J., Zhu, X., Messer, J. S., Murphy, S. F., Reddy, N., et al. (2011). "TNFAIP3 maintains intestinal barrier function and supports epithelial cell tight junctions". *PLoS One* 6: e26352. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0026352.

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). "Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9". *Nat Biotechnol* 34: pp. 192–8. ISSN: 1087-0156. DOI: 10.1038/nbt.3450.

Kozomara, A. and Griffiths-Jones, S. (2014). "miRBase: annotating high confidence microRNAs using deep sequencing data". *Nucleic Acids Res* 42: pp. D68–73. ISSN: 0305-1048. DOI: 10.1093/nar/gkt1181.

Kruger, P., Saffarzadeh, M., Weber, A. N., Rieber, N., Radsak, M., Bernuth, H. von, Benarafa, C., Roos, D., Skokowa, J., and Hartl, D. (2015). "Neutrophils: Between host defence, immune modulation, and tissue injury". *PLoS Pathog* 11: e1004651. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.1004651.

Ku, C. S., Loy, E. Y., Salim, A., Pawitan, Y., and Chia, K. S. (2010). "The discovery of human genetic variations and their use as disease markers: past, present and future". *J Hum Genet* 55: pp. 403–15. ISSN: 1434-5161. DOI: 10.1038/jhg.2010.55.

Kucuk, A., Uslu, A. U., Ugan, Y., Bagcaci, S., Karahan, A. Y., Akarmut, A., Sahin, A., and Kucuksen, S. (2015). "Neutrophil-to-lymphocyte ratio is involved in the severity of ankylosing spondylitis". *Bratisl Lek Listy* 116: pp. 722–5. ISSN: 0006-9248 (Print) 0006-9248.

Kumar, V., Westra, H. J., Karjalainen, J., Zhernakova, D. V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., Vosa, U., et al. (2013). "Human disease-associated genetic variation impacts large intergenic non-coding RNA expression". *PLoS Genet* 9: e1003201. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1003201.

Lai, N. S., Yu, H. C., Chen, H. C., Yu, C. L., Huang, H. B., and Lu, M. C. (2013). "Aberrant expression of microRNAs in T cells from patients with ankylosing spondylitis contributes to the immunopathogenesis". *Clinical & Experimental Immunology* 173: pp. 47–57. ISSN: 1365-2249. DOI: 10.1111/cei.12089. URL: http://dx.doi.org/10.1111/cei.12089.

Lamot, L., Borovecki, F., Tambic Bukovac, L., Vidovic, M., Perica, M., Gotovac, K., and Harjacek, M. (2014). "Aberrant expression of shared master-key genes contributes to the immunopathogenesis in patients with juvenile spondyloarthritis". *PLoS One* 9: e115416. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0115416.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., et al. (2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia". *Genome Res* 22: pp. 1813–31. ISSN: 1088-9051. DOI: 10.1101/gr.136184.111.

Lange, K. M. de, Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S. G., et al. (2017). "Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease". *Nat Genet* 49: pp. 256–261. ISSN: 1061-4036. DOI: 10.1038/ng.3760.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". *Genome Biol* 10: R25. ISSN: 1474-7596. DOI: 10.1186/gb-2009-10-3-r25.

Langrish, C. L., Chen, Y., Blumenschein, W. M., Mattson, J., Basham, B., Sedgwick, J. D., McClanahan, T., Kastelein, R. A., and Cua, D. J. (2005). "IL-23 drives a pathogenic T cell population that induces autoimmune inflammation". *J Exp Med* 201: pp. 233–40. ISSN: 0022-1007 (Print) 0022-1007. DOI: 10.1084/jem.20041257.

Laval, S. H., Timms, A., Edwards, S., Bradbury, L., Brophy, S., Milicic, A., Rubin, L., Siminovitch, K. A., Weeks, D. E., Calin, A., et al. (2001). "Whole-Genome Screening in Ankylosing Spondylitis: Evidence of Non-MHC Genetic-Susceptibility Loci". *Am J Hum Genet* 68: pp. 918–26. ISSN: 0002-9297 (Print) 1537-6605 (Electronic).

Lawrence, T., Gilroy, D. W., Colville-Nash, P. R., and Willoughby, D. A. (2001). "Possible new role for NF-kappaB in the resolution of inflammation". *Nat Med* 7: pp. 1291–7. ISSN: 1078-8956 (Print) 1078-8956. DOI: 10.1038/nm1201-1291.

Lawrence, T., Bebien, M., Liu, G. Y., Nizet, V., and Karin, M. (2005). "IKKalpha limits macrophage NF-kappaB activation and contributes to the resolution of inflammation". *Nature* 434: pp. 1138–43. ISSN: 0028-0836. DOI: 10.1038/nature03491.

Leirisalo-Repo, Marjatta, Hannu, Timo, and Mattila, Leena (2003). "Microbial factors in spondyloarthropathies: insights from population studies". *Curr Opin Rheumatol* 15: pp. 408–412. ISSN: 1040-8711. URL: http://journals.lww.com/co-rheumatology/Fulltext/2003/07000/Microbial_factors_in_spondyloarthropathies_.6.aspx.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). "The Sequence Alignment/Map format and SAMtools". *Bioinformatics* 25: pp. 2078–9. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp352.

Li, Mulin J., Wang, Lily Y., Xia, Zhengyuan, Sham, Pak C., and Wang, Junwen (2013). "GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications". *Nucleic acids research* 41: p. 8. ISSN: 0305-1048. DOI: 10.1093/nar/gkt456. URL: http://dx.doi.org/10.1093/nar/gkt456.

Li, P., Zheng, Y., and Chen, X. (2017a). "Drugs for Autoimmune Inflammatory Diseases: From Small Molecule Compounds to Anti-TNF Biologics". *Front Pharmacol* 8: p. 460. ISSN: 1663-9812 (Print) 1663-9812. DOI: 10.3389/fphar.2017.00460.

Li, S., Miao, T., Sebastian, M., Bhullar, P., Ghaffari, E., Liu, M., Symonds, A. L., and Wang, P. (2012). "The transcription factors Egr2 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells". *Immunity* 37: pp. 685–96. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2012.08.001.

Li, S., Tighe, S. W., Nicolet, C. M., Grove, D., Levy, S., Farmerie, W., Viale, A., Wright, C., Schweitzer, P. A., Gao, Y., et al. (2014a). "Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study". *Nat Biotechnol* 32: pp. 915–925. ISSN: 1087-0156. DOI: 10.1038/nbt.2972.

Li, X., Chai, W., Zhang, G., Ni, M., Chen, J., Dong, J., Zhou, Y., Hao, L., Bai, Y., and Wang, Y. (2017b). "Down-Regulation of lncRNA-AK001085 and its Influences on the Diagnosis of Ankylosing Spondylitis". *Med Sci Monit* 23: pp. 11–16. ISSN: 1234-1010.

Li, X. L., Yu, H., and Wu, G. S. (2014b). "Investigating the genetic association of HCP5, SPATA2, TNIP1, TNFAIP3 and COG6 with psoriasis in Chinese population". *Int J Immunogenet* 41: pp. 503–7. ISSN: 1744-3121. DOI: 10.1111/iji.12150.

Li, Y. and Tollefsbol, T. O. (2011). "DNA methylation detection: bisulfite genomic sequencing analysis". *Methods Mol Biol* 791: pp. 11–21. ISSN: 1064-3745. DOI: 10.1007/978-1-61779-316-5_2.

Li, Z., Wong, S. H., Shen, J., Chan, M. T., and Wu, W. K. (2016). "The Role of MicroRNAS in Ankylosing Spondylitis". *Medicine (Baltimore)* 95: e3325. ISSN: 0025-7974. DOI: 10.1097/md.0000000000003325.

Li, Z., Haynes, K., Pennisi, D. J., Anderson, L. K., Song, X., Thomas, G. P., Kenna, T., Leo, P., and Brown, M. A. (2017c). "Epigenetic and gene expression analysis of ankylosing spondylitis-associated loci implicate immune cells and the gut in the disease pathogenesis". *Genes Immun*. ISSN: 1466-4879. DOI: 10.1038/gene.2017.11.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs". *Nature* 433: pp. 769–73. ISSN: 0028-0836. DOI: 10.1038/nature03315.

Lin, P., Bach, M., Asquith, M., Lee, A. Y., Akileswaran, L., Stauffer, P., Davin, S., Pan, Y., Cambronne, E. D., Dorris, M., et al. (2014). "HLA-B27 and human beta2-microglobulin affect the gut microbiota of transgenic rats". *PLoS One* 9: e105684. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0105684.

Lippiello, L., Yamamoto, K., Robinson, D., and Mankin, H. J. (1978). "Involvement of prostaglandins from rheumatoid synovium in inhibition of articular cartilage metabolism". *Arthritis Rheum* 21: pp. 909–17. ISSN: 0004-3591 (Print) 0004-3591.

Lipton, Sarah and Deodhar, Atul (2012). *The new ASAS classification criteria for axial and peripheral spondyloarthritis: Promises and pitfalls*. Vol. 7: pp. 675–682. DOI: 10.2217/ijr.12.61.

Liu, Jimmy Z., Sommeren, Suzanne van, Huang, Hailiang, Ng, Siew C., Alberts, Rudi, Takahashi, Atsushi, Ripke, Stephan, Lee, James C., Jostins, Luke, Shah, Tejas, et al. (2015). "Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations". *Nature genetics*. ISSN: 1061-4036. DOI: 10.1038/ng.3359. URL: http://dx.doi.org/10.1038/ng.3359.

Liu, Y., Ferguson, J. F., Xue, C., Silverman, I. M., Gregory, B., Reilly, M. P., and Li, M. (2013). "Evaluating the impact of sequencing depth on transcriptome profiling in human adipose". *PLoS One* 8: e66883. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0066883.

Liu, Y., Ye, Z., Li, X., Anderson, J. L., Khan, M., DaSilva, D., Baron, M., Wilson, D., Bocoun, V., Ivacic, L. C., et al. (2017). "Genetic and Functional Associations with Decreased Anti-inflammatory Tumor Necrosis Factor Alpha Induced Protein 3 in Macrophages from Subjects with Axial Spondyloarthritis". *Front Immunol* 8: p. 860. ISSN: 1664-3224 (Print) 1664-3224. DOI: 10.3389/fimmu.2017.00860.

Lofgren, S. E., Frostegard, J., Truedsson, L., Pons-Estel, B. A., D'Alfonso, S., Witte, T., Lauwerys, B. R., Endreffy, E., Kovacs, L., Vasconcelos, C., et al. (2012). "Genetic association of miRNA-146a with systemic lupus erythematosus in Europeans through decreased expression of the gene". *Genes Immun* 13: pp. 268–274. ISSN: 1466-4879. URL: http://dx.doi.org/10.1038/gene.2011.84.

Lopez-Larrea, C., Sujirachato, K., Mehra, N. K., Chiewsilp, P., Isarangkura, D., Kanga, U., Dominguez, O., Coto, E., Pena, M., Setien, F., et al. (1995). "HLA-B27 subtypes in Asian patients with ankylosing spondylitis. Evidence for new associations". *Tissue Antigens* 45: pp. 169–76. ISSN: 0001-2815 (Print) 0001-2815.

Lories, R. J. and Baeten, D. L. (2009). "Differences in pathophysiology between rheumatoid arthritis and ankylosing spondylitis". *Clin Exp Rheumatol* 27: S10–4. ISSN: 0392-856X (Print) 0392-856x.

Love, Michael I., Huber, Wolfgang, and Anders, Simon (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biology* 15: p. 550. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8. URL: https://doi.org/10.1186/s13059-014-0550-8http://dx.doi.org/10.1186/s13059-014-0550-8.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., et al. (2005). "MicroRNA expression profiles classify human cancers". *Nature* 435: pp. 834–8. ISSN: 0028-0836. DOI: 10.1038/nature03702.

Lubrano, Ennio and Perrotta, Fabio Massimo (2016). "Secukinumab for ankylosing spondylitis and psoriatic arthritis". *Therapeutics and Clinical Risk Management* 12: pp. 1587–1592. ISSN: 1178-203X. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5085310/.

Lukas, C., Landewe, R., Sieper, J., Dougados, M., Davis, J., Braun, J., Linden, S. van der, and Heijde, D. van der (2009). "Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis". *Ann Rheum Dis* 68: pp. 18–24. ISSN: 0003-4967. DOI: 10.1136/ard.2008.094870.

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., et al. (2015). "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions". *Cell* 161: pp. 1012–25. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.04.004.

Lv, F., Huang, Y., Lv, W., Yang, L., Li, F., Fan, J., and Sun, J. (2017). "MicroRNA-146a Ameliorates Inflammation via TRAF6/NF-kappaB Pathway in Intervertebral Disc Cells". *Med Sci Monit* 23: pp. 659–664. ISSN: 1234-1010.

Ma, A. and Malynn, B. A. (2012). "A20: linking a complex regulator of ubiquitylation to immunity and human disease". *Nat Rev Immunol* 12: pp. 774–85. ISSN: 1474-1733. DOI: 10.1038/nri3313.

Ma, H., Ye, B., Wei, Q., and Zhu, X. D. (2013). "Genome wide gene expression analysis of macrophages from ankylosing spondylitis patients under interferon-gamma treatment". *Eur Rev Med Pharmacol Sci* 17: pp. 2798–803. ISSN: 1128-3602 (Print) 1128-3602.

Ma, X., Becker Buscaglia, L. E., Barker, J. R., and Li, Y. (2011). "MicroRNAs in NF-kappaB signaling". *J Mol Cell Biol* 3: pp. 159–66. ISSN: 1759-4685. DOI: 10.1093/jmcb/mjr007.

MacKay, K., Mack, C., Brophy, S., and Calin, A. (1998). "The Bath Ankylosing Spondylitis Radiology Index (BASRI): a new, validated approach to disease assessment". *Arthritis Rheum* 41: pp. 2263–70. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/1529-0131(199812)41:12<2263::aid-art23>3.0.co;2-i.

MacLean, I. L., Iqball, S., Woo, P., Keat, A. C., Hughes, R. A., Kingsley, G. H., and Knight, S. C. (1993). "HLA-B27 subtypes in the spondarthropathies". *Clinical and Experimental Immunology* 91: pp. 214–219. ISSN: 0009-9104 1365-2249. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1554677/.

Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., Horikoshi, M., Johnson, A. D., Ng, M. C., Prokopenko, I., et al. (2014). "Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility". *Nat Genet* 46: pp. 234–44. ISSN: 1061-4036. DOI: 10.1038/ng.2897.

Maher, B. (2008). "Personal genomes: The case of the missing heritability". *Nature* 456: pp. 18–21. ISSN: 0028-0836. DOI: 10.1038/456018a.

Maksymowych, W. P., Morency, N., Conner-Spady, B., and Lambert, R. G. (2013). "Suppression of inflammation and effects on new bone formation in ankylosing spondylitis: evidence for a window of opportunity in disease modification". *Ann Rheum Dis* 72: pp. 23–8. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2011-200859.

Makunin, I. V., Pheasant, M., Simons, C., and Mattick, J. S. (2007). "Orthologous microRNA genes are located in cancer-associated genomic regions in human and mouse". *PLoS One* 2: e1133. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0001133.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). "RNA-guided human genome engineering via Cas9". *Science* 339: pp. 823–6. ISSN: 0036-8075. DOI: 10.1126/science.1232033.

Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J. M., Auton, A., Myers, S., Morris, A., et al. (2012). "Bayesian refinement of association signals for 14 loci in 3 common diseases". *Nat Genet* 44: pp. 1294–301. ISSN: 1061-4036. DOI: 10.1038/ng.2435.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., et al. (2009). "Finding the missing heritability of complex diseases". *Nature* 461: pp. 747–53. ISSN: 0028-0836. DOI: 10.1038/nature08494.

Marchini, J. and Howie, B. (2010). "Genotype imputation for genome-wide association studies". *Nat Rev Genet* 11: pp. 499–511. ISSN: 1471-0056. DOI: 10.1038/nrg2796.

Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). "A new multipoint method for genome-wide association studies by imputation of genotypes". *Nat Genet* 39. DOI: 10.1038/ng2088. URL: https://doi.org/10.1038/ng2088.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). "RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays". *Genome Res* 18: pp. 1509–17. ISSN: 1088-9051 (Print) 1088-9051. DOI: 10.1101/gr.079558.108.

Marques, R. E., Guabiraba, R., Russo, R. C., and Teixeira, M. M. (2013). "Targeting CCL5 in inflammation". *Expert Opin Ther Targets* 17: pp. 1439–60. ISSN: 1472-8222. DOI: 10.1517/14728222.2013.837886.

Martin, Marcel (2011). *Cutadapt removes adapter sequences from high-throughput sequencing reads*. Vol. 17. DOI: 10.14806/ej.17.1.200.

Masters, S. L., Simon, A., Aksentijevich, I., and Kastner, D. L. (2009). "Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (*)". *Annu Rev Immunol* 27: pp. 621–68. ISSN: 0732-0582 (Print) 0732-0582. DOI: 10.1146/annurev.immunol.25.022106.141627.

Maston, Glenn A., Evans, Sara K., and Green, Michael R. (2006). "Transcriptional Regulatory Elements in the Human Genome". *Genomics and Human Genetics* 7: pp. 29–59. ISSN: 1527-8204. DOI: 10.1146/annurev.genom.7.080505.115623. URL: http://dx.doi.org/10.1146/annurev.genom.7.080505.115623.

Matmati, M., Jacques, P., Maelfait, J., Verheugen, E., Kool, M., Sze, M., Geboes, L., Louagie, E., Mc Guire, C., Vereecke, L., et al. (2011). "A20 (TNFAIP3) deficiency in myeloid cells triggers erosive polyarthritis resembling rheumatoid arthritis". *Nat Genet* 43: pp. 908–12. ISSN: 1061-4036. DOI: 10.1038/ng.874.

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). "Systematic localization of common disease-associated variation in regulatory DNA". *Science* 337: pp. 1190–5. ISSN: 0036-8075. DOI: 10.1126/science.1222794.

Maurano, Matthew T., Haugen, Eric, Sandstrom, Richard, Vierstra, Jeff, Shafer, Anthony, Kaul, Rajinder, and Stamatoyannopoulos, John A. (2015). "Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo". *Nature genetics* 47: pp. 1393–1401. ISSN: 1061-4036. DOI: 10.1038/ng.3432. URL: http://dx.doi.org/10.1038/ng.3432.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges". *Nat Rev Genet* 9: pp. 356–69. ISSN: 1471-0056. DOI: 10.1038/nrg2344.

McClay, J. L., Shabalin, A. A., Dozmorov, M. G., Adkins, D. E., Kumar, G., Nerella, S., Clark, S. L., Bergen, S. E., Hultman, C. M., Magnusson, P. K., et al. (2015). "High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction". *Genome Biol* 16: p. 291. ISSN: 1474-7596. DOI: 10.1186/s13059-015-0842-7.

McCoy, J. M., Wicks, J. R., and Audoly, L. P. (2002). "The role of prostaglandin E2 receptors in the pathogenesis of rheumatoid arthritis". *J Clin Invest* 110: pp. 651–8. ISSN: 0021-9738 (Print) 0021-9738. DOI: 10.1172/jci15528.

McGovern, Amanda, Schoenfelder, Stefan, Martin, Paul, Massey, Jonathan, Duffus, Kate, Plant, Darren, Yarwood, Annie, Pratt, Arthur G., Anderson, Amy E., Isaacs, John D., et al. (2016). "Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23". *Genome biology* 17: p. 212. ISSN: 1474-7596. DOI: 10.1186/s13059-016-1078-x. URL: http://dx.doi.org/10.1186/s13059-016-1078-x.

McGovern, D. P., Gardet, A., Torkvist, L., Goyette, P., Essers, J., and Taylor, K. D. (2010). "Genome-wide association identifies multiple ulcerative colitis susceptibility loci". *Nat Genet* 42. DOI: 10.1038/ng.549. URL: https://doi.org/10.1038/ng.549.

Mercan, R., Bitik, B., Tufan, A., Bozbulut, U. B., Atas, N., Ozturk, M. A., Haznedaroglu, S., and Goker, B. (2016). "The Association Between Neutrophil/Lymphocyte Ratio and Disease Activity in Rheumatoid Arthritis and Ankylosing Spondylitis". *J Clin Lab Anal* 30: pp. 597–601. ISSN: 0887-8013. DOI: 10.1002/jcla.21908.

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). "Long non-coding RNAs: insights into functions". *Nat Rev Genet* 10: pp. 155–9. ISSN: 1471-0056. DOI: 10.1038/nrg2521.

Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J., and Alba, M. M. (2002). "PROMO: detection of known transcription regulatory elements using species-tailored searches". *Bioinformatics* 18: pp. 333–4. ISSN: 1367-4803 (Print) 1367-4803.

Miceli-Richard, C., Zouali, H., Said-Nahal, R., Lesage, S., Merlin, F., De Toma, C., Blanche, H., Sahbatou, M., Dougados, M., Thomas, G., et al. (2004). "Significant linkage to spondyloarthropathy on 9q31-34". *Hum Mol Genet* 13: pp. 1641–8. ISSN: 0964-6906 (Print) 0964-6906. DOI: 10.1093/hmg/ddh179.

Mielants, H., Veys, E. M., Cuvelier, C., De Vos, M., Goemaere, S., De Clercq, L., Schatteman, L., and Elewaut, D. (1995). "The evolution of spondyloarthropathies in relation to gut histology. II. Histological aspects". *J Rheumatol* 22: pp. 2273–8. ISSN: 0315-162X (Print) 0315-162X.

Mifsud, Borbala, Tavares-Cadete, Filipe, Young, Alice N., Sugar, Robert, Schoenfelder, Stefan, Ferreira, Lauren, Wingett, Steven W., Andrews, Simon, Grey, William, Ewels, Philip A., et al. (2015). "Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C". *Nature genetics*. ISSN: 1061-4036. DOI: 10.1038/ng.3286. URL: http://dx.doi.org/10.1038/ng.3286.

Migliore, A., Broccoli, S., Bizzi, E., and Lagana, B. (2012). "Indirect comparison of the effects of anti-TNF biological agents in patients with ankylosing spondylitis by means of a mixed treatment comparison performed on efficacy data from published randomised, controlled trials". *J Med Econ* 15: pp. 473–80. ISSN: 1369-6998. DOI: 10.3111/13696998.2012.660255.

Milanez, Fernanda M., Saad, Carla G., Viana, Vilma T., Moraes, Jlio C. C., Prico, Grgory V. V., Sampaio-Barros, Percival D., Goncalves, Clio R. R., and Bonf, Elosa (2015). "IL-23/Th17 axis is not influenced by TNF-blocking agents in ankylosing spondylitis patients". *Arthritis research & therapy* 18: p. 52. ISSN: 1478-6354. DOI: 10.1186/s13075-016-0949-6. URL: http://dx.doi.org/10.1186/s13075-016-0949-6.

Mitoma, H., Horiuchi, T., Tsukamoto, H., and Ueda, N. (2016). "Molecular mechanisms of action of anti-TNF-alpha agents - Comparison among therapeutic TNF-alpha antagonists". *Cytokine*. ISSN: 1043-4666. DOI: 10.1016/j.cyto.2016.08.014.

Monaco, Claudia, Nanchahal, Jagdeep, Taylor, Peter, and Feldmann, Marc (2015). "Anti-TNF therapy: past, present and future". *International Immunology* 27: pp. 55–62. ISSN: 0953-8178 1460-2377. DOI: 10.1093/intimm/dxu102. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4279876/.

Montilla, C., Del Pino-Montes, J., Collantes-Estevez, E., Font, P., Zarco, P., Mulero, J., Gratacos, J., Rodriguez, C., Juanola, X., Fernandez-Sueiro, J. L., et al. (2012). "Clinical features of late-onset ankylosing spondylitis: comparison with early-onset disease". *J Rheumatol* 39: pp. 1008–12. ISSN: 0315-162X (Print) 0315-162X. DOI: 10.3899/jrheum.111082.

Moratz, C., Harrison, K., and Kehrl, J. H. (2004). "Regulation of chemokine-induced lymphocyte migration by RGS proteins". *Methods Enzymol* 389: pp. 15–32. ISSN: 0076-6879 (Print) 0076-6879. DOI: 10.1016/s0076-6879(04)89002-5.

Morris, G. E. (2008). "The Cajal body". *Biochim Biophys Acta* 1783: pp. 2108–15. ISSN: 0006-3002 (Print) 0006-3002. DOI: 10.1016/j.bbamcr.2008.07.016.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq". *Nat Methods* 5: pp. 621–8. ISSN: 1548-7091. DOI: 10.1038/nmeth.1226.

Mourtada-Maarabouni, M., Hasan, A. M., Farzaneh, F., and Williams, G. T. (2010). "Inhibition of human T-cell proliferation by mammalian target of rapamycin (mTOR) antagonists requires noncoding RNA growth-arrest-specific transcript 5 (GAS5)". *Mol Pharmacol* 78: pp. 19–28. ISSN: 0026-895x. DOI: 10.1124/mol.110.064055.

Muller, N., Doring, F., Klapper, M., Neumann, K., Schulte, D. M., Turk, K., Schroder, J. O., Zeuner, R. A., Freitag-Wolf, S., Schreiber, S., et al. (2014). "Interleukin-6 and tumour necrosis factor-alpha differentially regulate lincRNA transcripts in cells of the innate immune system in vivo in human subjects with rheumatoid arthritis". *Cytokine* 68: pp. 65–8. ISSN: 1043-4666. DOI: 10.1016/j.cyto.2014.03.004.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., Li, X., Li, H., Kuperwasser, N., Ruda, V. M., et al. (2010). "From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus". *Nature* 466: pp. 714–9. ISSN: 0028-0836. DOI: 10.1038/nature09266.

Myers, R. A., Scott, N. M., Gauderman, W. J., Qiu, W., Mathias, R. A., Romieu, I., Levin, A. M., Pino-Yanes, M., Graves, P. E., Villarreal, A. B., et al. (2014). "Genome-wide interaction studies reveal sex-specific asthma risk alleles". *Hum Mol Genet* 23: pp. 5251–9. ISSN: 0964-6906. DOI: 10.1093/hmg/ddu222.

Nakasa, T., Miyaki, S., Okubo, A., Hashimoto, M., Nishida, K., Ochi, M., and Asahara, H. (2008). "Expression of microRNA-146 in rheumatoid arthritis synovial tissue". *Arthritis Rheum* 58: pp. 1284–92. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/art.23429.

Nenci, A., Becker, C., Wullaert, A., Gareus, R., Loo, G. van, Danese, S., Huth, M., Nikolaev, A., Neufert, C., Madison, B., et al. (2007). "Epithelial NEMO links innate immunity to chronic intestinal inflammation". *Nature* 446: pp. 557–61. ISSN: 0028-0836. DOI: 10.1038/nature05698.

Niu, Z., Wang, J., Zou, H., Yang, C., Huang, W., and Jin, L. (2015). "Common MIR146A Polymorphisms in Chinese Ankylosing Spondylitis Subjects and Controls". *PLoS One* 10: e0137770. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0137770.

Oberst, A., Dillon, C. P., Weinlich, R., McCormick, L. L., Fitzgerald, P., Pop, C., Hakem, R., Salvesen, G. S., and Green, D. R. (2011). "Catalytic activity of the caspase-8-FLIP(L) complex inhibits RIPK3-dependent necrosis". *Nature* 471: pp. 363–7. ISSN: 0028-0836. DOI: 10.1038/nature09852.

Odhams, Christopher A., Cortini, Andrea, Chen, Lingyan, Roberts, Amy L., Viuela, Ana, Buil, Alfonso, Small, Kerrin S., Dermitzakis, Emmanouil T., Morris, David L., Vyse, Timothy J., et al. (2017). "Mapping eQTLs with RNA-seq reveals novel susceptibility genes, non-coding RNAs and alternative-splicing events in systemic lupus erythematosus". *Human molecular genetics*. ISSN: 0964-6906. DOI: 10.1093/hmg/ddw417. URL: http://dx.doi.org/10.1093/hmg/ddw417.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., et al. (2001). "A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease". *Nature* 411: pp. 603–6. ISSN: 0028-0836 (Print) 0028-0836. DOI: 10.1038/35079114.

Oliver, J., Rueda, B., Lopez-Nevot, M. A., Gomez-Garcia, M., and Martin, J. (2007). "Replication of an association between IL23R gene polymorphism with inflammatory bowel disease". *Clin Gastroenterol Hepatol* 5: 977–81, 981.e1–2. ISSN: 1542-3565. DOI: 10.1016/j.cgh.2007.05.002.

Olsson, A. H., Volkov, P., Bacos, K., Dayeh, T., Hall, E., Nilsson, E. A., Ladenvall, C., Ronn, T., and Ling, C. (2014). "Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets". *PLoS Genet* 10: e1004735. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1004735.

Orange, J. S. and Ballas, Z. K. (2006). "Natural killer cells in human health and disease". *Clin Immunol* 118: pp. 1–10. ISSN: 1521-6616 (Print) 1521-6616. DOI: 10.1016/j.clim.2005.10.011.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). "Latent enhancers activated by stimulation in differentiated cells". *Cell* 152: pp. 157–71. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.12.018.

Ozen, Gulsen, Deniz, Rabia, Eren, Fatih, Erzik, Can, Unal, Ali Ugur, Yavuz, Sule, Aydin, Sibel Zehra, Inanc, Nevsun, Direskeneli, Haner, and Atagunduz, Pamir (2017). "Association of ERAP1, IL23R and PTGER4 Polymorphisms with Radiographic Severity of Ankylosing Spondylitis". *The Open Rheumatology Journal* 11: pp. 1–9. ISSN: 1874-3129. DOI: 10.2174/1874312901711010001. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5366379/.

Paladini, F., Belfiore, F., Cocco, E., Carcassi, C., Cauli, A., Vacca, A., Fiorillo, M. T., Mathieu, A., Cascino, I., and Sorrentino, R. (2009). "HLA-E gene polymorphism associates with ankylosing spondylitis in Sardinia". *Arthritis Res Ther* 11: R171. ISSN: 1478-6354. DOI: 10.1186/ar2860.

Pang, K. C., Dinger, M. E., Mercer, T. R., Malquori, L., Grimmond, S. M., Chen, W., and Mattick, J. S. (2009). "Genome-wide identification of long noncoding RNAs in CD8+ T cells". *J Immunol* 182: pp. 7738–48. ISSN: 0022-1767. DOI: 10.4049/jimmunol.0900603.

Papanagnou, Panagiota, Stivarou, Theodora, and Tsironi, Maria (2016). "The Role of miRNAs in Common Inflammatory Arthropathies: Osteoarthritis and Gouty Arthritis". *Biomolecules* 6. ISSN: 2218-273X. DOI: 10.3390/biom6040044. URL: http://dx.doi.org/10.3390/biom6040044.

Paraskevi, A., Theodoropoulos, G., Papaconstantinou, I., Mantzaris, G., Nikiteas, N., and Gazouli, M. (2012). "Circulating MicroRNA in inflammatory bowel disease". *J Crohns Colitis* 6: pp. 900–4. ISSN: 1873-9946. DOI: 10.1016/j.crohns.2012.02.006.

Park, H., Li, Z., Yang, X. O., Chang, S. H., Nurieva, R., Wang, Y. H., Wang, Y., Hood, L., Zhu, Z., Tian, Q., et al. (2005a). "A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17". *Nat Immunol* 6: pp. 1133–41. ISSN: 1529-2908 (Print) 1529-2908. DOI: 10.1038/ni1261.

Park, J. M., Greten, F. R., Wong, A., Westrick, R. J., Arthur, J. S., Otsu, K., Hoffmann, A., Montminy, M., and Karin, M. (2005b). "Signaling pathways and genes that inhibit pathogen-induced macrophage apoptosis–CREB and NF-kappaB as key regulators". *Immunity* 23: pp. 319–29. ISSN: 1074-7613 (Print) 1074-7613. DOI: 10.1016/j.immuni.2005.08.010.

Park, S. G. and Kim, S. (2006). "Do aminoacyl-tRNA synthetases have biological functions other than in protein biosynthesis?" *IUBMB Life* 58: pp. 556–8. ISSN: 1521-6543 (Print) 1521-6543. DOI: 10.1080/15216540600735974.

Park, S. G., Schimmel, P., and Kim, S. (2008). "Aminoacyl tRNA synthetases and their connections to disease". *Proc Natl Acad Sci U S A* 105: pp. 11043–9. ISSN: 0027-8424. DOI: 10.1073/pnas.0802862105.

Park, Y. W., Kee, S. J., Cho, Y. N., Lee, E. H., Lee, H. Y., Kim, E. M., Shin, M. H., Park, J. J., Kim, T. J., Lee, S. S., et al. (2009). "Impaired differentiation and cytotoxicity of natural killer cells in systemic lupus erythematosus". *Arthritis Rheum* 60: pp. 1753–63. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/art.24556.

Parker, Stephen C., Stitzel, Michael L., Taylor, D. L., Orozco, Jose M., Erdos, Michael R., Akiyama, Jennifer A., Bueren, Kelly L. van, Chines, Peter S., Narisu, Narisu, Program, Nisc, et al. (2013). "Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants". *Proceedings of the National Academy of Sciences of the United States of America* 110: pp. 17921–17926. ISSN: 0027-8424. DOI: 10.1073/pnas.1317023110. URL: http://dx.doi.org/10.1073/pnas.1317023110.

Parkes, Miles, Cortes, Adrian, Heel, David A. van, and Brown, Matthew A. (2013). "Genetic insights into common pathways and complex relationships among immune-mediated diseases". *Nat Rev Genet* 14: pp. 661–673. ISSN: 1471-0056. DOI: 10.1038/nrg3502http://www.nature.com/nrg/journal/v14/n9/abs/nrg3502.html#supplementary-information. URL: http://dx.doi.org/10.1038/nrg3502.

Paul, D. S. and Beck, S. (2014). "Advances in epigenome-wide association studies for common diseases". *Trends Mol Med* 20: pp. 541–3. ISSN: 1471-4914. DOI: 10.1016/j.molmed.2014.07.002.

Pauley, K. M., Satoh, M., Chan, A. L., Bubb, M. R., Reeves, W. H., and Chan, E. K. (2008). "Upregulated miR-146a expression in peripheral blood mononuclear cells from rheumatoid arthritis patients". *Arthritis Res Ther* 10: R101. ISSN: 1478-6354. DOI: 10.1186/ar2493.

Pelletier, M., Maggi, L., Micheletti, A., Lazzeri, E., Tamassia, N., Costantini, C., Cosmi, L., Lunardi, C., Annunziato, F., Romagnani, S., et al. (2010). "Evidence for a cross-talk between human neutrophils and Th17 cells". *Blood* 115: pp. 335–43. ISSN: 0006-4971. DOI: 10.1182/blood-2009-04-216085.

Peng, X., Gralinski, L., Armour, C. D., Ferris, M. T., Thomas, M. J., Proll, S., Bradel-Tretheway, B. G., Korth, M. J., Castle, J. C., Biery, M. C., et al. (2010). "Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling". *MBio* 1. DOI: 10.1128/mBio.00206-10.

Perbal, B. (2006). "New insight into CCN3 interactions–nuclear CCN3 : fact or fantasy?" *Cell Commun Signal* 4: p. 6. ISSN: 1478-811x. DOI: 10.1186/1478-811x-4-6.

Perl, A. (2016). "Activation of mTOR (mechanistic target of rapamycin) in rheumatic diseases". *Nat Rev Rheumatol* 12: pp. 169–82. ISSN: 1759-4790. DOI: 10.1038/nrrheum.2015.172.

Pickrell, J. K. (2014). "Joint analysis of functional genomic data and genome-wide association studies of 18 human traits". *Am J Hum Genet* 94: pp. 559–73. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2014.03.004.

Pimentel-Santos, F. M., Ligeiro, D., and Matos, M. (2011). "Whole blood transcriptional profiling in ankylosing spondylitis identifies novel candidate genes that might contribute to the inflammatory and tissue-destructive ". *Whole blood transcriptional profiling in ankylosing spondylitis identifies novel candidate genes that might contribute to the inflammatory and tissue-destructive*. URL: http://www.biomedcentral.com/content/pdf/ar3309.pdf.

Poddubnyy, D. and Rudwaleit, M. (2012). "Early spondyloarthritis". *Rheum Dis Clin North Am* 38: pp. 387–403. ISSN: 1558-3163 (Electronic) 0889-857X (Linking). DOI: 10.1016/j.rdc.2012.04.007.

Podolsky, D. K. (1997). "Lessons from genetic models of inflammatory bowel disease". *Acta Gastroenterol Belg* 60: pp. 163–5. ISSN: 0001-5644 (Print) 0001-5644.

Pointon, J. J., Harvey, D., Karaderi, T., Appleton, L. H., Farrar, C., Stone, M. A., Sturrock, R. D., Brown, M. A., and Wordsworth, B. P. (2010). "Elucidating the chromosome 9 association with AS; CARD9 is a candidate gene". *Genes Immun* 11: pp. 490–6. ISSN: 1476-5470 (Electronic) 1466-4879 (Linking). DOI: 10.1038/gene.2010.17.

Portanova, J. P., Zhang, Y., Anderson, G. D., Hauser, S. D., Masferrer, J. L., Seibert, K., Gregory, S. A., and Isakson, P. C. (1996). "Selective neutralization of prostaglandin E2 blocks inflammation, hyperalgesia, and interleukin 6 production in vivo". *J Exp Med* 184: pp. 883–91. ISSN: 0022-1007 (Print) 0022-1007.

Qian, B. P., Ji, M. L., Qiu, Y., Wang, B., Yu, Y., Shi, W., and Luo, Y. F. (2016). "Identification of Serum miR-146a and miR-155 as Novel Noninvasive Complementary Biomarkers for Ankylosing Spondylitis". *Spine (Phila Pa 1976)* 41: pp. 735–42. ISSN: 0362-2436. DOI: 10.1097/brs.0000000000001339.

Qu, K., Zaba, L. C., Satpathy, A. T., Giresi, P. G., Li, R., Jin, Y., Armstrong, R., Jin, C., Schmitt, N., Rahbar, Z., et al. (2017). "Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors". *Cancer Cell* 32: 27–41.e4. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2017.05.008.

Rashid, N. U., Giresi, P. G., Ibrahim, J. G., Sun, W., and Lieb, J. D. (2011). "ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions". *Genome Biol* 12: R67. ISSN: 1474-7596. DOI: 10.1186/gb-2011-12-7-r67.

Rendeiro, A. F., Schmidl, C., Strefford, J. C., Walewska, R., Davis, Z., Farlik, M., Oscier, D., and Bock, C. (2016). "Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks". *Nat Commun* 7: p. 11938. ISSN: 2041-1723. DOI: 10.1038/ncomms11938.

Reveille, J. D. (2006). "Major histocompatibility genes and ankylosing spondylitis". *Best Pract Res Clin Rheumatol* 20: pp. 601–9. ISSN: 1521-6942 (Print) 1521-6942 (Linking). DOI: 10.1016/j.berh.2006.03.004.

Reveille, J. D. (2010). "Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci". *Nat Genet* 42: pp. 123–127. ISSN: 1061-4036. DOI: http://www.nature.com/ng/journal/v42/n2/suppinfo/ng.513_S1.html. URL: http://dx.doi.org/10.1038/ng.513.

Reveille, J. D. (2011a). "Epidemiology of spondyloarthritis in North America". *Am J Med Sci* 341: pp. 284–6. ISSN: 0002-9629. DOI: 10.1097/MAJ.0b013e31820f8c99.

Reveille, J. D. (2011b). "The genetic basis of spondyloarthritis". *Ann Rheum Dis* 70 Suppl 1: pp. i44–50. ISSN: 1468-2060 (Electronic) 0003-4967 (Linking). DOI: 10.1136/ard.2010.140574.

Reveille, J. D. (2013). "A registry of ankylosing spondylitis registries and prospects for global interfacing". *Curr Opin Rheumatol* 25: pp. 468–76. ISSN: 1531-6963 (Electronic) 1040-8711 (Linking). DOI: 10.1097/BOR.0b013e3283620e1d.

Revell, P. A. and Mayston, V. (1982). "Histopathology of the synovial membrane of peripheral joints in ankylosing spondylitis". *Ann Rheum Dis* 41: pp. 579–86. ISSN: 0003-4967 (Print) 0003-4967.

Ritchlin, C. T., Kavanaugh, A., Gladman, D. D., Mease, P. J., Helliwell, P., Boehncke, W. H., Vlam, K. de, Fiorentino, D., Fitzgerald, O., Gottlieb, A. B., et al. (2009). "Treatment recommendations for psoriatic arthritis". *Ann Rheum Dis* 68: pp. 1387–94. ISSN: 0003-4967. DOI: 10.1136/ard.2008.094946.

Rittore, Ccile, Sanchez, Elodie, Soler, Stephan, Barat-Houari, Mouna, Albers, Marieke, Obici, Laura, McDermott, Michael F, Touitou, Isabelle, and Grandemange, Sylvie (2014). "Identification of a new exon 2-skipped TNFR1 transcript: regulation by three functional polymorphisms of the TNFR-associated periodic syndrome (TRAPS) gene". *Annals of the Rheumatic Diseases* 73: pp. 290–297. DOI: 10.1136/annrheumdis-2012-203023. URL: http://ard.bmj.com/content/73/1/290.abstract.

Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., et al. (2011). "Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease". *Nat Genet* 43: pp. 1066–73. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.952.

Roberts, Amity R., Vecellio, Matteo, Chen, Liye, Ridley, Anna, Cortes, Adrian, Knight, Julian C., Bowness, Paul, Cohen, Carla J., and Wordsworth, B. P. (2016). "An ankylosing spondylitis-associated genetic variant in the IL23R-IL12RB2 intergenic region modulates enhancer activity and is associated with increased Th1-cell differentiation". *Annals of the rheumatic diseases*. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2015-208640. URL: http://dx.doi.org/10.1136/annrheumdis-2015-208640.

Roberts, Amity R., Appleton, Louise H., Cortes, Adrian, Vecellio, Matteo, Lau, Jonathan, Watts, Laura, Brown, Matthew A., and Wordsworth, Paul (2017). "ERAP1 association with ankylosing spondylitis is attributable to common genotypes rather than rare haplotype combinations". *Proceedings of the National Academy of Sciences of the United States of America*. ISSN: 0027-8424. DOI: 10.1073/pnas.1618856114. URL: http://dx.doi.org/10.1073/pnas.1618856114.

Robinson, D. R., Tashjian A. H., Jr., and Levine, L. (1975). "Prostaglandin-stimulated bone resorption by rheumatoid synovia. A possible mechanism for bone destruction in rheumatoid arthritis". *J Clin Invest* 56: pp. 1181–8. ISSN: 0021-9738 (Print) 0021-9738. DOI: 10.1172/jci108195.

Roda, G., Jharap, B., Neeraj, N., and Colombel, J. F. (2016). "Loss of Response to Anti-TNFs: Definition, Epidemiology, and Management". *Clin Transl Gastroenterol* 7: e135. ISSN: 2155-384X (Print). DOI: 10.1038/ctg.2015.63.

Rosenbaum, James T. and Davey, Michael P. (2011). "Time for a gut check: Evidence for the hypothesis that HLAB27 predisposes to ankylosing spondylitis by altering the microbiome". *Arthritis & Rheumatism* 63: pp. 3195–3198. ISSN: 1529-0131. DOI: 10.1002/art.30558. URL: http://dx.doi.org/10.1002/art.30558.

Round, June L. and Mazmanian, Sarkis K. (2009). "The gut microbiome shapes intestinal immune responses during health and disease". *Nature reviews. Immunology* 9: pp. 313–323. ISSN: 1474-1733 1474-1741. DOI: 10.1038/nri2515. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4095778/.

Rudwaleit, M. and Baeten, D. (2006). "Ankylosing spondylitis and bowel disease". *Best Pract Res Clin Rheumatol* 20: pp. 451–71. ISSN: 1521-6942 (Print) 1521-6942 (Linking). DOI: 10.1016/j.berh.2006.03.010.

Rudwaleit, M., Listing, J., Brandt, J., Braun, J., and Sieper, J. (2004). "Prediction of a major clinical response (BASDAI 50) to tumour necrosis factor alpha blockers in ankylosing spondylitis". *Ann Rheum Dis* 63: pp. 665–70. ISSN: 0003-4967 (Print) 0003-4967. DOI: 10.1136/ard.2003.016386.

Sand, M., Gambichler, T., Sand, D., Skrygan, M., Altmeyer, P., and Bechara, F. G. (2009). "MicroRNAs and the skin: tiny players in the body's largest organ". *J Dermatol Sci* 53: pp. 169–75. ISSN: 0923-1811 (Print) 0923-1811. DOI: 10.1016/j.jdermsci.2008.10.004.

Sandborn, W. J., Hanauer, S. B., Katz, S., Safdi, M., Wolf, D. G., Baerg, R. D., Tremaine, W. J., Johnson, T., Diehl, N. N., and Zinsmeister, A. R. (2001). "Etanercept for active Crohn's disease: a randomized, double-blind, placebo-controlled trial". *Gastroenterology* 121: pp. 1088–94. ISSN: 0016-5085 (Print) 0016-5085.

Sanjana, N. E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F. (2016). "High-resolution interrogation of functional elements in the noncoding genome". *Science* 353: pp. 1545–1549. ISSN: 0036-8075. DOI: 10.1126/science.aaf7613.

Santoro, M., Nociti, V., Lucchini, M., De Fino, C., Losavio, F. A., and Mirabella, M. (2016). "Expression Profile of Long Non-Coding RNAs in Serum of Patients with Multiple Sclerosis". *J Mol Neurosci* 59: pp. 18–23. ISSN: 0895-8696. DOI: 10.1007/s12031-016-0741-8.

Scharer, C. D., Blalock, E. L., Barwick, B. G., Haines, R. R., Wei, C., Sanz, I., and Boss, J. M. (2016). "ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naive SLE B cells". *Sci Rep* 6: p. 27030. ISSN: 2045-2322. DOI: 10.1038/srep27030.

Schlosstein, L., Terasaki, P. I., Bluestone, R., and Pearson, C. M. (1973). "High association of an HL-A antigen, W27, with ankylosing spondylitis". *N Engl J Med* 288: pp. 704–6. ISSN: 0028-4793 (Print) 0028-4793. DOI: 10.1056/nejm197304052881403.

Schofield, E. C., Carver, T., Achuthan, P., Freire-Pritchett, P., Spivakov, M., Todd, J. A., and Burren, O. S. (2016). "CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets". *Bioinformatics* 32: pp. 2511–3. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw173.

Schoors, S., De Bock, K., Cantelmo, A. R., Georgiadou, M., Ghesquiere, B., Cauwenberghs, S., Kuchnio, A., Wong, B. W., Quaegebeur, A., Goveia, J., et al. (2014). "Partial and transient reduction of glycolysis by PFKFB3 blockade reduces pathological angiogenesis". *Cell Metab* 19: pp. 37–48. ISSN: 1550-4131. DOI: 10.1016/j.cmet.2013.11.008.

Schork, N. J., Murray, S. S., Frazer, K. A., and Topol, E. J. (2009). "Common vs. rare allele hypotheses for complex diseases". *Curr Opin Genet Dev* 19: pp. 212–9. ISSN: 0959-437X. DOI: 10.1016/j.gde.2009.04.010.

Schreiber, E., Matthias, P., Muller, M. M., and Schaffner, W. (1989). "Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells". *Nucleic Acids Res* 17: p. 6419. ISSN: 0305-1048 (Print) 0305-1048.

Schumacher, M. A., Hedl, M., Abraham, C., Bernard, J. K., Lozano, P. R., Hsieh, J. J., Almohazey, D., Bucar, E. B., Punit, S., Dempsey, P. J., et al. (2017). "ErbB4 signaling stimulates pro-inflammatory macrophage apoptosis and limits colonic inflammation". *Cell Death Dis* 8: e2622. DOI: 10.1038/cddis.2017.42.

Sharaf, Nazar, Nicklin, Martin J., and Giovine, Francesco S. (2014). "Long-range DNA interactions at the IL-1/IL-36/IL-37 gene cluster (2q13) are induced by activation of monocytes". *Cytokine* 68: pp. 16–22. ISSN: 1043-4666. DOI: 10.1016/j.cyto.2014.03.002. URL: http://dx.doi.org/10.1016/j.cyto.2014.03.002.

Sharma, Choi, D., and Planck, S. R. (2009). "Insights in to the pathogenesis of axial spondyloarthropathy based on gene expression profiles". *Insights in to the pathogenesis of axial spondyloarthropathy based on gene expression profiles*. URL: http://www.biomedcentral.com/content/pdf/ar2855.pdfYou.

Shen, Hui, Goodall, Jane C., and Hill Gaston, J. S. (2009). "Frequency and phenotype of peripheral blood Th17 cells in ankylosing spondylitis and rheumatoid arthritis". *Arthritis and rheumatism* 60: pp. 1647–1656. ISSN: 0004-3591. DOI: 10.1002/art.24568. URL: http://dx.doi.org/10.1002/art.24568.

Shen, Li, Shao, Ning-Yi Y., Liu, Xiaochuan, Maze, Ian, Feng, Jian, and Nestler, Eric J. (2013). "diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates". *PloS one* 8. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0065598. URL: http://dx.doi.org/10.1371/journal.pone.0065598.

Shi, L., Zhang, Z., Yu, A. M., Wang, W., Wei, Z., Akhter, E., Maurer, K., Costa Reis, P., Song, L., Petri, M., et al. (2014). "The SLE transcriptome exhibits evidence of chronic endotoxin exposure and has widespread dysregulation of non-coding and coding RNAs". *PLoS One* 9: e93846. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0093846.

Shibayama, Y., Fanucchi, S., Magagula, L., and Mhlanga, M. M. (2014). "lncRNA and gene looping: whats the connection?" *Transcription* 5. ISSN: 2154-1264 (Print). DOI: 10.4161/trns.28658.

Sieper, J. and Poddubnyy, D. (2016). "New evidence on the management of spondyloarthritis". *Nat Rev Rheumatol* 12: pp. 282–95. ISSN: 1759-4790. DOI: 10.1038/nrrheum.2016.42.

Sigdel, K. R., Cheng, A., Wang, Y., Duan, L., and Zhang, Y. (2015). "The Emerging Functions of Long Noncoding RNA in Immune Cells: Autoimmune Diseases". *J Immunol Res* 2015: p. 848790. ISSN: 2314-7156. DOI: 10.1155/2015/848790.

Sims, A. M., Barnardo, M., Herzberg, I., Bradbury, L., Calin, A., Wordsworth, B. P., Darke, C., and Brown, M. A. (2007). "Non-B27 MHC associations of ankylosing spondylitis". *Genes Immun* 8: pp. 115–23. ISSN: 1466-4879 (Print) 1466-4879. DOI: 10.1038/sj.gene.6364362.

Smemo, S., Tena, J. J., Kim, K. H., Gamazon, E. R., and Sakabe, N. J. (2014). "Obesity-associated variants within FTO form long-range functional connections with IRX3". *Obesity-associated variants within FTO form long-range functional connections with IRX3*. ISSN: 0028-0836. DOI: 10.1038/nature13138. URL: http://dx.doi.org/10.1038/nature13138.

Smith, A. K., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., Tylavsky, F. A., and Conneely, K. N. (2014). "Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type". *BMC Genomics* 15: p. 145. ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-145.

Smith, J. A., Barnes, Hong, D., and DeLay, M. L. (2008). "Gene expression analysis of macrophages derived from ankylosing spondylitis patients reveals interferon dysregulation". *Gene expression analysis of macrophages derived from ankylosing spondylitis patients reveals interferon dysregulation*. ISSN: 1529-0131. DOI: 10.1002/art.23512. URL: http://dx.doi.org/10.1002/art.23512.

Smyth, D. J., Plagnol, V., Walker, N. M., Cooper, J. D., Downes, K., Yang, J. H., Howson, J. M., Stevens, H., McManus, R., Wijmenga, C., et al. (2008). "Shared and distinct genetic variants in type 1 diabetes and celiac disease". *N Engl J Med* 359: pp. 2767–77. ISSN: 0028-4793. DOI: 10.1056/NEJMoa0807917.

Soldner, F., Stelzer, Y., Shivalila, C. S., Abraham, B. J., Latourelle, J. C., Barrasa, M. I., Goldmann, J., Myers, R. H., Young, R. A., and Jaenisch, R. (2016). "Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression". *Nature* 533: pp. 95–9. ISSN: 0028-0836. DOI: 10.1038/nature17939.

Song, I. H., Appel, H., Haibel, H., Loddenkemper, C., Braun, J., Sieper, J., and Rudwaleit, M. (2008). "New onset of Crohn's disease during treatment of active ankylosing spondylitis with etanercept". *J Rheumatol* 35: pp. 532–6. ISSN: 0315-162X (Print) 0315-162x.

Song, J., Kim, D., Han, J., Kim, Y., Lee, M., and Jin, E. J. (2015). "PBMC and exosome-derived Hotair is a critical regulator and potent marker for rheumatoid arthritis". *Clin Exp Med* 15: pp. 121–6. ISSN: 1591-8890. DOI: 10.1007/s10238-013-0271-4.

Song, L. and Crawford, G. E. (2010). "DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells". *Cold Spring Harb Protoc* 2010. DOI: 10.1101/pdb.prot5384. URL: http://dx.doi.org/10.1101/pdb.prot5384.

Sonkoly, E. and Pivarcsi, A. (2009). "Advances in microRNAs: implications for immunity and inflammatory diseases". *J Cell Mol Med* 13: pp. 24–38. ISSN: 1582-1838. DOI: 10.1111/j.1582-4934.2008.00534.x.

Sonkoly, E., Bata-Csorgo, Z., Pivarcsi, A., Polyanka, H., Kenderessy-Szabo, A., Molnar, G., Szentpali, K., Bari, L., Megyeri, K., Mandi, Y., et al. (2005). "Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS". *J Biol Chem* 280: pp. 24159–67. ISSN: 0021-9258 (Print) 0021-9258. DOI: 10.1074/jbc.M501704200.

Sonkoly, E., Wei, T., Janson, P. C., Saaf, A., Lundeberg, L., Tengvall-Linder, M., Norstedt, G., Alenius, H., Homey, B., Scheynius, A., et al. (2007). "MicroRNAs: novel regulators involved in the pathogenesis of psoriasis?" *PLoS One* 2: e610. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000610.

Sonkoly, E., Stahle, M., and Pivarcsi, A. (2008). "MicroRNAs and immunity: novel players in the regulation of normal immune function and inflammation". *Semin Cancer Biol* 18: pp. 131–40. ISSN: 1044-579X. DOI: 10.1016/j.semcancer.2008.01.005.

Spadaro, A., Scrivo, R., Moretti, T., Bernardini, G., Riccieri, V., Taccari, E., Strom, R., and Valesini, G. (2004). "Natural killer cells and gamma/delta T cells in synovial fluid and in peripheral blood of patients with psoriatic arthritis". *Clin Exp Rheumatol* 22: pp. 389–94. ISSN: 0392-856X (Print) 0392-856X.

Spitz, F. and Furlong, E. E. (2012). "Transcription factors: from enhancer binding to developmental control". *Nat Rev Genet* 13: pp. 613–26. ISSN: 1471-0056. DOI: 10.1038/nrg3207.

Stanczyk, J., Pedrioli, D. M., Brentano, F., Sanchez-Pernaute, O., Kolling, C., Gay, R. E., Detmar, M., Gay, S., and Kyburz, D. (2008). "Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis". *Arthritis Rheum* 58: pp. 1001–9. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/art.23386.

Stepensky, D., Bangia, N., and Cresswell, P. (2007). "Aggregate formation by ERp57-deficient MHC class I peptide-loading complexes". *Traffic* 8: pp. 1530–42. ISSN: 1398-9219 (Print) 1398-9219. DOI: 10.1111/j.1600-0854.2007.00639.x.

Stolwijk, C., Tubergen, A. van, Castillo-Ortiz, J. D., and Boonen, A. (2015a). "Prevalence of extra-articular manifestations in patients with ankylosing spondylitis: a systematic review and meta-analysis". *Ann Rheum Dis* 74: pp. 65–73. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2013-203582.

Stolwijk, C., Essers, I., Tubergen, A. van, Boonen, A., Bazelier, M. T., De Bruin, M. L., and Vries, F. de (2015b). "The epidemiology of extra-articular manifestations in ankylosing spondylitis: a population-based matched cohort study". *Ann Rheum Dis* 74: pp. 1373–8. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2014-205253.

Stone, M. A., Payne, U., Pacheco-Tena, C., and Inman, R. D. (2004). "Cytokine correlates of clinical response patterns to infliximab treatment of ankylosing spondylitis". *Ann Rheum Dis* 63: pp. 84–7. ISSN: 0003-4967 (Print) 0003-4967.

Strange, A., Capon, F., Spencer, C. C., Knight, J., Weale, M. E., Allen, M. H., Barton, A., Band, G., Bellenguez, C., Bergboer, J. G., et al. (2010). "A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1". *Nat Genet* 42: pp. 985–90. ISSN: 1546-1718 (Electronic) 1061-4036 (Linking). DOI: 10.1038/ng.694.

Su, H. C. and Lenardo, M. J. (2008). "Genetic defects of apoptosis and primary immunodeficiency". *Immunol Allergy Clin North Am* 28: pp. 329–51, ix. ISSN: 0889-8561 (Print) 0889-8561. DOI: 10.1016/j.iac.2008.01.002.

Suarez-Gestal, M., Calaza, M., Endreffy, E., Pullmann, R., Ordi-Ros, J., Sebastiani, G. D., Ruzickova, S., Jose Santos, M., Papasteriades, C., Marchini, M., et al. (2009). "Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study". *Arthritis Res Ther* 11: R69. ISSN: 1478-6354. DOI: 10.1186/ar2698.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles". *Proc Natl Acad Sci U S A* 102: pp. 15545–50. ISSN: 0027-8424 (Print) 0027-8424. DOI: 10.1073/pnas.0506580102.

Sun, D., Yu, Z., Fang, X., Liu, M., Pu, Y., Shao, Q., Wang, D., Zhao, X., Huang, A., Xiang, Z., et al. (2017). "LncRNA GAS5 inhibits microglial M2 polarization and exacerbates demyelination". *EMBO Rep*. ISSN: 1469-221X. DOI: 10.15252/embr.201643668.

Szalay, B., Meszaros, G., Cseh, A., Acs, L., Deak, M., Kovacs, L., Vasarhelyi, B., and Balog, A. (2012). "Adaptive immunity in ankylosing spondylitis: phenotype and functional alterations of T-cells before and during infliximab therapy". *Clin Dev Immunol* 2012: p. 808724. ISSN: 1740-2522. DOI: 10.1155/2012/808724.

Taganov, K. D., Boldin, M. P., Chang, K. J., and Baltimore, D. (2006). "NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses". *Proc Natl Acad Sci U S A* 103: pp. 12481–6. ISSN: 0027-8424 (Print) 0027-8424. DOI: 10.1073/pnas.0605298103.

Takahasi, F., Shah, H. S., Wise, L. S., and Das, K. M. (1990). "Circulating antibodies against human colonic extract enriched with a 40 kDa protein in patients with ulcerative colitis". *Gut* 31: pp. 1016–20. ISSN: 0017-5749 (Print) 0017-5749.

Talpin, A., Costantino, F., Bonilla, N., Leboime, A., Letourneur, F., Jacques, S., Dumont, F., Amraoui, S., Dutertre, C. A., Garchon, H. J., et al. (2014). "Monocyte-derived dendritic cells from HLA-B27+ axial spondyloarthritis (SpA) patients display altered functional capacity and deregulated gene expression". *Arthritis Res Ther* 16: p. 417. ISSN: 1478-6354. DOI: 10.1186/s13075-014-0417-0.

Tamouza, R., Mansour, I., Bouguacha, N., Klayme, S., Djouadi, K., Laoussadi, S., Azoury, M., Dulphy, N., Ramasawmy, R., Krishnamoorthy, R., et al. (2001). "A new HLA-B*27 allele (B*2719) identified in a Lebanese patient affected with ankylosing spondylitis". *Tissue Antigens* 58: pp. 30–3. ISSN: 0001-2815 (Print) 0001-2815.

Tan, F. K. and Farheen, K. (2011). "The potential importance of Toll-like receptors in ankylosing spondylitis". *Int J Clin Rheumtol* 6: pp. 649–54. ISSN: 1758-4272 (Print) 1758-4280 (Electronic). DOI: 10.2217/ijr.11.61.

Tang, Y., Luo, X., Cui, H., Ni, X., Yuan, M., Guo, Y., Huang, X., Zhou, H., Vries, N. de, Tak, P. P., et al. (2009). "MicroRNA-146A contributes to abnormal activation of the type I interferon pathway in human lupus by targeting the key signaling proteins". *Arthritis Rheum* 60: pp. 1065–75. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/art.24436.

Tani, H., Onuma, Y., Ito, Y., and Torimura, M. (2014). "Long non-coding RNAs as surrogate indicators for chemical stress responses in human-induced pluripotent stem cells". *PLoS One* 9: e106282. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0106282.

Tavares, R. M., Turer, E. E., Liu, C. L., Advincula, R., Scapini, P., Rhee, L., Barrera, J., Lowell, C. A., Utz, P. J., Malynn, B. A., et al. (2010). "The ubiquitin modifying enzyme A20 restricts B cell survival and prevents autoimmunity". *Immunity* 33: pp. 181–91. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2010.07.017.

Thomas, G. P., Duan, R., Pettit, A. R., Weedon, H., Kaur, S., Smith, M., and Brown, M. A. (2013). "Expression profiling in spondyloarthropathy synovial biopsies highlights changes in expression of inflammatory genes in conjunction with tissue remodelling genes". *BMC Musculoskelet Disord*. Vol. 14: p. 354. ISBN: 1471-2474 (Electronic). DOI: 10.1186/1471-2474-14-354.

Thurston, T. L., Ryzhakov, G., Bloor, S., Muhlinen, N. von, and Randow, F. (2009). "The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria". *Nat Immunol* 10: pp. 1215–21. ISSN: 1529-2908. DOI: 10.1038/ni.1800.

Togayachi, A., Kozono, Y., Kuno, A., Ohkura, T., Sato, T., Hirabayashi, J., Ikehara, Y., and Narimatsu, H. (2010). "Beta3GnT2 (B3GNT2), a major polylactosamine synthase: analysis of B3GNT2-deficient mice". *Methods Enzymol* 479: pp. 185–204. ISSN: 0076-6879. DOI: 10.1016/s0076-6879(10)79011-x.

Toussirot, E., Abbas, W., Khan, K. A., Tissot, M., Jeudy, A., Baud, L., Bertolini, E., Wendling, D., and Herbein, G. (2013). "Imbalance between HAT and HDAC activities in the PBMCs of patients with ankylosing spondylitis or rheumatoid arthritis and influence of HDAC inhibitors on TNF alpha production". *PLoS One* 8: e70939. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0070939.

Trynka, G., Zhernakova, A., Romanos, J., Franke, L., Hunt, K. A., Turner, G., Bruinenberg, M., Heap, G. A., Platteel, M., Ryan, A. W., et al. (2009). "Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling". *Gut* 58: pp. 1078–83. ISSN: 0017-5749. DOI: 10.1136/gut.2008.169052.

Trynka, Gosia, Hunt, Karen A., Bockett, Nicholas A., Romanos, Jihane, Mistry, Vanisha, Szperl, Agata, Bakker, Sjoerd F., Bardella, Maria Teresa, Bhaw-Rosun, Leena, Castillejo, Gemma, et al. (2011). "Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease". *Nature genetics* 43: pp. 1193–1201. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3242065/.

Trynka, Gosia, Sandor, Cynthia, Han, Buhm, Xu, Han, Stranger, Barbara E., Liu, Shirley X., and Raychaudhuri, Soumya (2013). "Chromatin marks identify critical cell types for fine mapping complex trait variants". *Nature genetics* 45: pp. 124–130. DOI: 10.1038/ng.2504. URL: http://dx.doi.org/10.1038/ng.2504.

Tsoi, Lam C., Spain, Sarah L., Knight, Jo, Ellinghaus, Eva, Stuart, Philip E., Capon, Francesca, Ding, Jun, Li, Yanming, Tejasvi, Trilokraj, Gudjonsson, Johann E., et al. (2012). "Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity". *Nature genetics* 44: pp. 1341–1348. ISSN: 1061-4036. DOI: 10.1038/ng.2467. URL: http://dx.doi.org/10.1038/ng.2467.

Udalova, I. A., Knight, J. C., Vidal, V., Nedospasov, S. A., and Kwiatkowski, D. (1998). "Complex NF-kappaB interactions at the distal tumor necrosis factor promoter region in human monocytes". *J Biol Chem* 273: pp. 21178–86. ISSN: 0021-9258 (Print) 0021-9258.

Van Praet, L., Bosch, F. E. Van den, Jacques, P., Carron, P., Jans, L., Colman, R., Glorieus, E., Peeters, H., Mielants, H., De Vos, M., et al. (2013). "Microscopic gut inflammation in axial spondyloarthritis: a multiparametric predictive model". *Ann Rheum Dis* 72: pp. 414–7. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2012-202135.

Vecellio, M., Roberts, A. R., Cohen, C. J., and Cortes, A. (2015). "The genetic association of RUNX3 with ankylosing spondylitis can be explained by allele-specific effects on IRF4 recruitment that alter gene expression". *The genetic association of RUNX3 with ankylosing spondylitis can be explained by allele-specific effects on IRF4 recruitment that alter gene expression*. URL: https://ard.bmj.com/content/early/2015/10/09/annrheumdis-2015-207490.full.

Veer, L. J. van 't, Lutz, P. M., Isselbacher, K. J., and Bernards, R. (1992). "Structure and expression of major histocompatibility complex-binding protein 2, a 275-kDa zinc finger protein that binds to an enhancer of major histocompatibility complex class I genes". *Proceedings of the National Academy of Sciences of the United States of America* 89: pp. 8971–8975. ISSN: 0027-8424 1091-6490. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC50046/.

Veys, E. M. and Leare, M. van (1973). "Serum IgG, IgM, and IgA levels in ankylosing spondylitis". *Ann Rheum Dis* 32: pp. 493–6. ISSN: 0003-4967 (Print) 0003-4967.

Visscher, P. M. and Hill, W. G. (2009). "The limits of individual identification from sample allele frequencies: theory and statistical analysis". *PLoS Genet* 5: e1000628. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1000628.

Volinia, S., Calin, G. A., Liu, C. G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., et al. (2006). "A microRNA expression signature of human solid tumors defines cancer gene targets". *Proc Natl Acad Sci U S A* 103: pp. 2257–61. ISSN: 0027-8424 (Print) 0027-8424. DOI: 10.1073/pnas.0510565103.

Voswinkel, J., Weisgerber, K., Pfreundschuh, M., and Gause, A. (2001). "B lymphocyte involvement in ankylosing spondylitis: the heavy chain variable segment gene repertoire of B lymphocytes from germinal center-like foci in the synovial membrane indicates antigen selection". *Arthritis Res* 3: pp. 189–95. ISSN: 1465-9905 (Print) 1465-9905. DOI: 10.1186/ar297.

Vucic, D., Dixit, V. M., and Wertz, I. E. (2011). "Ubiquitylation in apoptosis: a post-translational modification at the edge of life and death". *Nat Rev Mol Cell Biol* 12: pp. 439–52. ISSN: 1471-0072. DOI: 10.1038/nrm3143.

Wakefield, J. (2007). "A Bayesian measure of the probability of false discovery in genetic epidemiology studies". *Am J Hum Genet* 81: pp. 208–27. ISSN: 0002-9297 (Print) 0002-9297. DOI: 10.1086/519024.

Wallace, C., Cutler, A. J., Pontikos, N., Pekalski, M. L., Burren, O. S., Cooper, J. D., Garcia, A. R., Ferreira, R. C., Guo, H., Walker, N. M., et al. (2015). "Dissection of a Complex Disease Susceptibility Region Using a Bayesian Stochastic Search Approach to Fine Mapping". *PLoS Genet* 11: e1005272. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1005272.

Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J. R., Xu, C., Futema, M., Lawson, D., et al. (2015). "The UK10K project identifies rare variants in health and disease". *Nature* 526: pp. 82–90. ISSN: 0028-0836. DOI: 10.1038/nature14962.

Wang, J., Zhuang, J., Iyer, S., Lin, X. Y., Greven, M. C., Kim, B. H., Moore, J., Pierce, B. G., Dong, X., Virgil, D., et al. (2013a). "Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium". *Nucleic Acids Res* 41: pp. D171–6. ISSN: 0305-1048. DOI: 10.1093/nar/gks1221.

Wang, J., Zeng, J., Wang, H., Ye, S., Bi, Y., Zhou, Y., and Li, K. (2016). "Genetic polymorphisms of RAGE and risk of ulcerative colitis in a Chinese population". *Immunol Lett* 170: pp. 88–94. ISSN: 0165-2478. DOI: 10.1016/j.imlet.2015.09.003.

Wang, J. H., Shih, K. S., Wu, Y. W., Wang, A. W., and Yang, C. R. (2013b). "Histone deacetylase inhibitors increase microRNA-146a expression and enhance negative regulation of interleukin-1beta signaling in osteoarthritis fibroblast-like synoviocytes". *Osteoarthritis Cartilage* 21: pp. 1987–96. ISSN: 1063-4584. DOI: 10.1016/j.joca.2013.09.008.

Wang, K. C. and Chang, H. Y. (2011). "Molecular mechanisms of long noncoding RNAs". *Mol Cell* 43: pp. 904–14. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2011.08.018.

Wang, Liuyang, Oehlers, Stefan H., Espenschied, Scott T., Rawls, John F., Tobin, David M., and Ko, Dennis C. (2015). "CPAG: software for leveraging pleiotropy in GWAS to reveal similarity between human traits links plasma fatty acids and intestinal inflammation". *Genome biology* 16: p. 190. ISSN: 1474-7596. DOI: 10.1186/s13059-015-0722-1. URL: http://dx.doi.org/10.1186/s13059-015-0722-1.

Wang, X. B., Ellis, J. J., Pennisi, D. J., Song, X., Batra, J., Hollis, K., Bradbury, L. A., Li, Z., Kenna, T. J., and Brown, M. A. (2017). "Transcriptome analysis of ankylosing spondylitis patients before and after TNF-[alpha] inhibitor therapy reveals the pathways affected". *Genes Immun*. ISSN: 1476-5470. DOI: 10.1038/gene.2017.19. URL: http://dx.doi.org/10.1038/gene.2017.19.

Wang, Z., Gerstein, M., and Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics". *Nat Rev Genet* 10: pp. 57–63. ISSN: 1471-0056. DOI: 10.1038/nrg2484.

Wapinski, O. and Chang, H. Y. (2011). "Long noncoding RNAs and human disease". *Trends Cell Biol* 21: pp. 354–61. ISSN: 0962-8924. DOI: 10.1016/j.tcb.2011.04.001.

Ward, L. D. and Kellis, M. (2012a). "Interpreting noncoding genetic variation in complex traits and human disease". *Nat Biotechnol* 30. DOI: 10.1038/nbt.2422. URL: https://doi.org/10.1038/nbt.2422.

Ward, Lucas D. and Kellis, Manolis (2012b). "HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants". *Nucleic Acids Research* 40: pp. D930–D934. ISSN: 0305-1048 1362-4962. DOI: 10.1093/nar/gkr917. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245002/.

Ward, M. M., Deodhar, A., Akl, E. A., Lui, A., Ermann, J., Gensler, L. S., Smith, J. A., Borenstein, D., Hiratzka, J., Weiss, P. F., et al. (2016). "American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network 2015 Recommendations for the Treatment of Ankylosing Spondylitis and Nonradiographic Axial Spondyloarthritis". *Arthritis Rheumatol* 68: pp. 282–98. ISSN: 2326-5191. DOI: 10.1002/art.39298.

Waterman, M., Xu, W., Stempak, J. M., Milgrom, R., Bernstein, C. N., Griffiths, A. M., Greenberg, G. R., Steinhart, A. H., and Silverberg, M. S. (2011). "Distinct and overlapping genetic loci in Crohn's disease and ulcerative colitis: correlations with pathogenesis". *Inflamm Bowel Dis* 17: pp. 1936–42. ISSN: 1078-0998. DOI: 10.1002/ibd.21579.

Werner, L., Guzner-Gur, H., and Dotan, I. (2013). "Involvement of CXCR4/CXCR7/CXCL12 Interactions in Inflammatory bowel disease". *Theranostics* 3: pp. 40–6. ISSN: 1838-7640. DOI: 10.7150/thno.5135.

White, M. A. (2015). "Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences". *Genomics* 106: pp. 165–70. ISSN: 0888-7543. DOI: 10.1016/j.ygeno.2015.06.003.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013). "Master transcription factors and mediator establish super-enhancers at key cell identity genes". *Cell* 153: pp. 307–19. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.03.035.

Wit, Elzo de and Laat, Wouter de (2012). "A decade of 3C technologies: insights into nuclear organization". *Genes & Development* 26: pp. 11–24. ISSN: 0890-9369. DOI: 10.1101/gad.179804.111. URL: http://dx.doi.org/10.1101/gad.179804.111.

Wit, Elzo de, Vos, Erica S., Holwerda, Sjoerd J., Valdes-Quezada, Christian, Verstegen, Marjon J., Teunissen, Hans, Splinter, Erik, Wijchers, Patrick J., Krijger, Peter H., and Laat, Wouter de (2015). "CTCF Binding Polarity Determines Chromatin Looping". *Molecular cell* 60: pp. 676–684. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.09.023. URL: http://dx.doi.org/10.1016/j.molcel.2015.09.023.

Wong, L., Jiang, K., Chen, Y., Hennon, T., Holmes, L., Wallace, C. A., and Jarvis, J. N. (2016). "Limits of Peripheral Blood Mononuclear Cells for Gene Expression-Based Biomarkers in Juvenile Idiopathic Arthritis". *Sci Rep* 6: p. 29477. ISSN: 2045-2322. DOI: 10.1038/srep29477.

Wong-Baeza, Isabel, Ridley, Anna, Shaw, Jackie, Hatano, Hiroko, Rysnik, Oliwia, McHugh, Kirsty, Piper, Christopher, Brackenbridge, Simon, Fernandes, Ricardo, Chan, Anthoni, et al. (2013). "KIR3DL2 Binds to HLA-B27 Dimers and Free H Chains More Strongly than Other HLA Class I and Promotes the Expansion of T Cells in Ankylosing Spondylitis". *The Journal of Immunology* 190: pp. 3216–3224. DOI: 10.4049/jimmunol.1202926. URL: http://www.jimmunol.org/content/190/7/3216.abstract.

Woo, M., Hakem, R., Furlonger, C., Hakem, A., Duncan, G. S., Sasaki, T., Bouchard, D., Lu, L., Wu, G. E., Paige, C. J., et al. (2003). "Caspase-3 regulates cell cycle in B cells: a consequence of substrate specificity". *Nat Immunol* 4: pp. 1016–22. ISSN: 1529-2908 (Print) 1529-2908. DOI: 10.1038/ni976.

Wright, C., Edelmann, M., diGleria, K., Kollnberger, S., Kramer, H., McGowan, S., McHugh, K., Taylor, S., Kessler, B., and Bowness, P. (2009). "Ankylosing spondylitis monocytes show upregulation of proteins involved in inflammation and the ubiquitin proteasome pathway". *Annals of the Rheumatic Diseases* 68: pp. 1626–1632. DOI: 10.1136/ard.2008.097204. URL: http://dx.doi.org/10.1136/ard.2008.097204.

WTCCC (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls". *Nature* 447: pp. 661–678. ISSN: 0028-0836. DOI: http://www.nature.com/nature/journal/v447/n7145/suppinfo/nature05911_S1.html. URL: http://dx.doi.org/10.1038/nature05911.

Wu, F., Zikusoka, M., Trindade, A., Dassopoulos, T., Harris, M. L., Bayless, T. M., Brant, S. R., Chakravarti, S., and Kwon, J. H. (2008). "MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha". *Gastroenterology* 135: 1624–1635.e24. ISSN: 0016-5085. DOI: 10.1053/j.gastro.2008.07.068.

Wu, Y., Wang, S., Farooq, S. M., Castelvetere, M. P., Hou, Y., Gao, J. L., Navarro, J. V., Oupicky, D., Sun, F., and Li, C. (2012). "A chemokine receptor CXCR2 macromolecular complex regulates neutrophil functions in inflammatory diseases". *J Biol Chem* 287: pp. 5744–55. ISSN: 0021-9258. DOI: 10.1074/jbc.M111.315762.

Wu, Y., Waite, L. L., Jackson, A. U., Sheu, W. H., Buyske, S., Absher, D., Arnett, D. K., Boerwinkle, E., Bonnycastle, L. L., Carty, C. L., et al. (2013). "Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained". *PLoS Genet* 9: e1003379. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.1003379.

Xie, Z., Li, J., Wang, P., Li, Y., Wu, X., Wang, S., Su, H., Deng, W., Liu, Z., Cen, S., et al. (2016). "Differential Expression Profiles of Long Noncoding RNA and mRNA of Osteogenically Differentiated Mesenchymal Stem Cells in Ankylosing Spondylitis". *J Rheumatol* 43: pp. 1523–31. ISSN: 0315-162X (Print) 0315-162x. DOI: 10.3899/jrheum.151181.

Xu, Hui, Wang, Zhang, Chen, Jing, Wang, Tian, Wang, Ling, Tang, Li, Lin, Xian-yang, Zhang, Chun-wu, and Chen, Bi-cheng (2015). "Association between Ankylosing Spondylitis and the miR-146a and miR-499 Polymorphisms". *PLOS ONE* 10. DOI: 10.1371/journal.pone.0122055. URL: http://dx.doi.org/10.1371/journal.pone.0122055.

Xue, H., Wu, J., Li, S., Rao, M. S., and Liu, Y. (2016). "Genetic Modification in Human Pluripotent Stem Cells by Homologous Recombination and CRISPR/Cas9 System". *Methods Mol Biol* 1307: pp. 173–90. ISSN: 1064-3745. DOI: 10.1007/7651_2014_73.

Yang, Y., Ma, Y., Shi, C., Chen, H., Zhang, H., Chen, N., Zhang, P., Wang, F., Yang, J., Yang, J., et al. (2013). "Overexpression of miR-21 in patients with ulcerative colitis impairs intestinal epithelial barrier function through targeting the Rho GTPase RhoB". *Biochem Biophys Res Commun* 434: pp. 746–52. ISSN: 0006-291x. DOI: 10.1016/j.bbrc.2013.03.122.

Yant, S. R., Zhu, W., Millinoff, D., Slightom, J. L., Goodman, M., and Gumucio, D. L. (1995). "High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster". *Nucleic Acids Research* 23: pp. 4353–4362. ISSN: 0305-1048 1362-4962. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC307390/.

Yu, P., Shen, F., Zhang, X., Cao, R., Zhao, X., Liu, P., Tu, H., Yang, X., Shi, R., and Zhang, H. (2012). "Association of single nucleotide polymorphisms of IL23R and IL17 with ulcerative colitis risk in a Chinese Han population". *PLoS One* 7: e44380. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0044380.

Yuan, F., Zhang, Y. H., Kong, X. Y., and Cai, Y. D. (2017). "Identification of Candidate Genes Related to Inflammatory Bowel Disease Using Minimum Redundancy Maximum Relevance, Incremental Feature Selection, and the Shortest-Path Approach". *Biomed Res Int* 2017: p. 5741948. DOI: 10.1155/2017/5741948.

Zhang, F., Wu, L., Qian, J., Qu, B., Xia, S., La, T., Wu, Y., Ma, J., Zeng, J., Guo, Q., et al. (2016). "Identification of the long noncoding RNA NEAT1 as a novel inflammatory regulator acting through MAPK pathway in human lupus". *J Autoimmun* 75: pp. 96–104. ISSN: 0896-8411. DOI: 10.1016/j.jaut.2016.07.012.

Zhang, G., Luo, J., Bruckel, J., Weisman, M. A., Schumacher, H. R., Khan, M. A., Inman, R. D., Mahowald, M., Maksymowych, W. P., Martin, T. M., et al. (2004). "Genetic studies in familial ankylosing spondylitis susceptibility". *Arthritis Rheum* 50: pp. 2246–54. ISSN: 0004-3591 (Print) 0004-3591. DOI: 10.1002/art.20308.

Zhang, X., Weissman, S. M., and Newburger, P. E. (2014a). "Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells". *RNA Biol* 11: pp. 777–87. ISSN: 1547-6286.

Zhang, X., Wang, C., Zhao, J., Xu, J., Geng, Y., Dai, L., Huang, Y., Fu, S. C., Dai, K., and Zhang, X. (2017). "miR-146a facilitates osteoarthritis by regulating cartilage homeostasis via targeting Camk2d and Ppp3r2". *Cell Death Dis* 8: e2734. DOI: 10.1038/cddis.2017.146.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). "Model-based analysis of ChIP-Seq (MACS)". *Genome Biol* 9: R137. ISSN: 1474-7596. DOI: 10.1186/gb-2008-9-9-r137.

Zhang, Z. H., Jhaveri, D. J., Marshall, V. M., Bauer, D. C., Edson, J., Narayanan, R. K., Robinson, G. J., Lundberg, A. E., Bartlett, P. F., Wray, N. R., et al. (2014b). "A comparative study of techniques for differential expression analysis on RNA-Seq data". *PLoS One* 9: e103207. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0103207.

Zheng, H. F., Forgetta, V., Hsu, Y. H., Estrada, K., Rosello-Diez, A., Leo, P. J., Dahia, C. L., Park-Min, K. H., Tobias, J. H., Kooperberg, C., et al. (2015). "Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture". *Nature* 526: pp. 112–7. ISSN: 0028-0836. DOI: 10.1038/nature14878.

Zhernakova, A., Festen, E. M., Franke, L., Trynka, G., Diemen, C. C. van, Monsuur, A. J., Bevova, M., Nijmeijer, R. M., Slot, R. van 't, Heijmans, R., et al. (2008). "Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP". *Am J Hum Genet* 82: pp. 1202–10. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2008.03.016.

Zhou, Q., Haupt, S., Kreuzer, J. T., Hammitzsch, A., Proft, F., Neumann, C., Leipe, J., Witt, M., Schulze-Koops, H., and Skapenko, A. (2015). "Decreased expression of miR-146a and miR-155 contributes to an abnormal Treg phenotype in patients with rheumatoid arthritis". *Ann Rheum Dis* 74: pp. 1265–74. ISSN: 0003-4967. DOI: 10.1136/annrheumdis-2013-204377.

Zochling, J., Heijde, D. van der, Dougados, M., and Braun, J. (2006). "Current evidence for the management of ankylosing spondylitis: a systematic literature review for the ASAS/EULAR management recommendations in ankylosing spondylitis". *Ann Rheum Dis* 65: pp. 423–32. ISSN: 0003-4967 (Print) 0003-4967. DOI: 10.1136/ard.2005.041129.

Zollner, Thomas M. and Asadullah, Khusru (2003). "Selectin and selectin ligand binding: a bittersweet attraction". *Journal of Clinical Investigation* 112: p. 980. DOI: 10.1172/jci19962. URL: http://dx.doi.org/10.1172/jci19962.

Zvyagin, Ivan V., Mamedov, Ilgar Z., Britanova, Olga V., Staroverov, Dmitriy B., Nasonov, Evgeni L., Bochkova, Anna G., Chkalina, Anna V., Kotlobay, Alexei A., Korostin, Dmitriy O., Rebrikov, Denis V., et al. (2010). "Contribution of functional KIR3DL1 to ankylosing spondylitis". *Cell Mol Immunol* 7: pp. 471–476. ISSN: 1672-7681. DOI: http://www.nature.com/cmi/journal/v7/n6/suppinfo/cmi201042s1.html. URL: http://dx.doi.org/10.1038/cmi.2010.42.