The genomics and genetics of ankylosing spondylitis

Abstract: The spondyloarthropathies are a group of arthritides which specifically target the spine and pelvis with ankylosing spondylitis (AS) being the most prevalent and debilitating of these conditions. Unique to AS is the progression to excessive uncontrolled bone formation following an initial inflammatory phase that can result in joint fusion and significant disability. Spondyloarthritis is estimated to affect 1%–2% of the population, twice as many as rheumatoid arthritis and thus constitutes a significant health problem. Currently AS pathogenesis is very poorly understood but recent large-scale genetics and gene expression profiling studies have identified some of the underlying mechanisms and pathways contributing to the disease. Genome-wide association studies have identified a number of candidate genes associated with AS sharing the same pathways which are now being targeted for therapeutic intervention. However, although such approaches can identify genes contributing to the disease process and are very informative as to disease aetiopathogenesis, they cannot profile the actual changes in gene/cell activity at any point in the disease process or possibly more importantly at specific sites. Such information is generated using expression profiling. A number of expression profiling studies have been undertaken in AS, looking at both circulating cells and tissues from affected joints. Although some common genes/pathways have been identified, overall the results to date have been somewhat disappointing due to differences in experimental design and tissue source as well as the low power of the studies. More recent better powered studies have shown some potential in developing gene expression profiling as a diagnostic tool in AS. True future success will rely on larger genetic and genomic studies and the combination of these datasets in eQTL studies requiring significant collaborative efforts. Such larger-scale approaches will also generate sufficient power to target specific disease stages and sites.

Keywords: ankylosing spondylitis, genomewide association studies, gene expression, microarrays, spondylarthritides, inflammatory arthritis

Introduction

The spondyloarthropathies are estimated to affect 1%–2% of the population almost two-fold higher prevalence than rheumatoid arthritis (RA). AS is the prototypic and most prevalent spondyloarthropathy and affects between 0.1%–1.4% of the global population. The condition primarily affects the spine and sacroiliac joints of the pelvis, causing pain and stiffness and eventual fusion. Asymmetric peripheral joint arthritis (~20%) and episodic acute anterior uveitis (~40%) are also common and approximately 10% of AS cases also have either inflammatory bowel disease (IBD) or psoriasis, while approximately 70% have sub-clinical gut inflammation. The disease onset is typically in the second decade of life and thus the disease has a higher impact relative to later onset diseases, such as rheumatoid arthritis, due to its longer disease duration.
Clinical features

The clinical features of AS can be divided into axial and peripheral components. Axial features include spondylitis, sacroiliitis and skeletal complications. Peripheral features include enthesitis, peripheral arthritis, inflammatory bowel disease (IBD) and anterior uveitis.

Axial features

Axial disease most commonly presents as inflammatory lower back pain and morning stiffness. Morning stiffness lasts from one to several hours and is improved by exercise but not relieved by rest (as in the case of mechanical back pain). Lower back pain results from inflammation of the SI joints and the vertebrae. Sacroiliitis, inflammation of the SI joint, is the most common feature of AS. Sacroiliitis can be detected by conventional x-ray of the pelvis which shows joint space narrowing and eventual sclerosis of the joint (Figure 1).

A distinguishing feature of AS is the site of the initial inflammation. This occurs at the entheses, the sites where tendons or ligaments join bone resulting in enthesitis one of the earliest features of AS (Figure 2). The stooped posture characteristic of AS patients is caused by inflammation of the spine and formation of syndesmophytes, bone growths originating from the initial sites of inflammation. In the spine of AS patients syndesmophytes form in the intervertebral joints and lead to fusion of the vertebrae which sometimes results in a classical ‘bamboo spine’ appearance. Fusion of bone across intervertebral joints ultimately causes flattening of the lumbar spine and limits neck motion and chest expansion. Paradoxically for a disease that induces new bone formation AS patients frequently exhibit reduced bone mineral density leading to osteoporosis driven by the high levels of inflammatory molecules associated with the disease.

Peripheral features

Arthritis of peripheral joints including hips, shoulders, knees, wrists, ankles and elbows is common in AS. Radiographically the peripheral arthritis in AS is similar to that observed in rheumatoid arthritis (RA). However, in AS additional ankylosis of wrists, hips, tarsal joints and small joints in the fingers and toes is common. Peripheral joint arthritis is usually asymmetric.

Extra-articular features

IBD

A close relationship exists between gut inflammation and AS. Approximately two-thirds of AS cases display signs of gut inflammation unrelated to clinical gastrointestinal symptoms, and AS cases and their first-degree relatives have increased gut permeability. As will be discussed in more detail later many genetic associations are shared between AS and both Crohn’s disease and ulcerative colitis, including IL-23R, STAT3, CARD9, IL12B, PTGER4, and KIF21B. These findings suggest common pathogenic processes shared between IBD and AS. Reactive arthritis, an AS- related spondyloarthritis, can result from acute gastrointestinal infections, and these findings suggest that the gut is a major source of antigenic drive in AS.
However the association between gut and joint disease is not fully understood. Activated T cells expressing members of the β7 family of integrins, which are involved in homing of intestinal lymphocytes, are enriched in the synovium of Spondyloarthropathy patients suggesting an intestinal origin to these cells.8

Anterior uveitis
Acute anterior uveitis (AAU) affects approximately 40% of patients with AS and approximately half of all AS patients with AU are B27+.9 AAU is normally unilateral and characterised by red, sore eyes, increased tear production and photophobia. AAU does not precede axial or peripheral inflammation but is often the first clinical sign presented to doctors by patients who may have lived with lower back pain for several years.

Cardiovascular disease
Recently it has been reported that AS patients are at increased risk for many cardiovascular disorders compared with the general population.10 The cardiovascular diseases include aortic valvular heart disease, non-aortic valvular heart disease, ischemic heart disease and congestive heart failure. Risk of developing these cardiovascular diseases is greatest in younger AS patients.10 The reasons for increased risk of cardiovascular events in AS patients are unclear. While it is difficult to determine the contribution of disease to mortality rates a recent study compared mortality among almost 700 AS patients and age, sex and locality-matched controls and found increased mortality in the AS group with circulatory disease the most frequent cause of death among the AS cases.11

Aetiology and pathology
There is a strong genetic component in the risk of developing AS. Studies in twins have estimated heritability of AS at >90%.12 HLA-B27 is established as the major genetic risk factor and the association between HLA-B27 and AS is among the strongest genetic association with any common disease. The genetic aetiology of AS will be discussed in greater detail below. Males are more frequently affected than females at a ratio of 2.5:1. The reasons for this imbalance are unclear.

The initial trigger for AS is currently unknown. With few exceptions, AS has a worldwide distribution, indicating that whatever the trigger it must be ubiquitous. Furthermore, epidemics of AS have not been reported. Studies in the B27 transgenic rat model of AS support the idea of a ubiquitous environmental trigger; B27 transgenic rats exposed to normal commensal bacteria develop disease while rats maintained in germ-free conditions do not.13 It has been suggested that this ubiquitous environmental factor may be a common microbial pathogen.14–16 A genetic defect reducing the immune system’s ability to clear the pathogen could result in a low-level sub-clinical infection which may result in the autoimmune response seen through the pathogen presenting self-antigens or the pathogen’s antigens mimicking self-antigens. Immune cells express various molecules capable of sensing a wide variety of microbial peptides. These include toll-like receptors (TLR), Dectin-1, PAMPs and DAMPs. Defects in recognition of microbial products by immune cells may contribute to an inflammatory cascade, driven by inflammatory cytokines such as IL-23, IL-1 and IL-6. The relative contributions of innate and adaptive immunity to the pathogenesis of AS is still unclear. Understanding the precise role of different immune cell types during different stages of disease and at different sites of disease is important for understanding how to develop novel targeted therapies. Recent studies in our lab have demonstrated that over 25% of γδ T cells in the peripheral blood of AS patients with active disease express IL-23R, secrete large amounts of IL-17 and display a strong Th17 bias.17 This data suggests an important role for γδ T cells in pathogenesis of AS. Given that γδ T cells bridge the innate-adaptive systems expressing myriad activation receptors such as Dectin-1, Toll-like receptors and NKG2D as well as IL-23R and IL-1, they are attractive candidate cells for further investigation in the pathogenesis of AS.

AS has two phases, an initial inflammatory phase followed by an osteoproliferative/ankylosing phase where the joint fusion is initiated and progresses. The inflammatory phase has similarities with other inflammatory arthrites such as RA with high levels of pro-inflammatory cytokine production and joint damage through osteoclast activity. However, the extent of joint destruction though the cytokine/osteoclast axis is significantly less than in RA with the dominant joint impact stemming from ankylosis occurring as a result of osteoproliferation. The most common site of this excessive bone formation in AS is in the cervical and lumbar vertebrae resulting in squaring of the vertebrae and eventually formation of syndesmophytes from the vertebral corners which can eventually bridge leading to ankylosis. Whether the initial inflammation directly links to the subsequent osteoproliferative stages is subject to significant debate. There is little evidence for a direct link between the two with some evidence that they are temporally distinct.18,19
This is supported by studies of the MRI evidence of inflammation and the subsequent development of syndesmophytes. These show that whilst there is an increased rate of syndesmophytes formation at vertebral corners with preceding MRI evidence of inflammation, most syndesmophytes arise at corners with no prior evidence of inflammation, and most corners with evidence of inflammation do not go on to develop syndesmophytes. These studies are consistent with models in which inflammation precedes bone formation but in which the link between inflammation and bone formation is not as direct as, for example, the link between inflammation and erosion in RA. There are therefore two key triggers in AS that are currently unknown; the signal for disease initiation, and the switch from inflammation to the damaging osteoproliferation phase. Elucidation of these triggers would be a major advance and identify new therapeutic approaches.

**Treatment**

The goal of treatment within AS is two-fold: to reduce pain and disability and to prevent progression of disease. Current therapies have provided symptom relief and reduced disability but have shown no efficacy in preventing or even slowing progression of disease. A key issue is the difficulty in diagnosing AS. Currently, firm diagnosis of AS requires confirmation of joint deterioration by X-ray (Modified New York Criteria), which only occurs in quite advanced disease, leading to an average delay between onset of symptoms and diagnosis of 8–11 years. Early diagnosis of AS would allow treatment to be commenced before significant joint damage has occurred possibly increasing the potential to slow disease progress. Trials investigating the effectiveness of this approach are currently underway.

Unlike patients with rheumatoid arthritis, AS patients respond well to physical therapy and long periods of immobility increase pain and immobility. Supervised group physiotherapy shows best symptom relief compared with group exercise or home exercise programs.

Conventional anti-rheumatic drugs such as methotrexate and sulfasalazine are ineffective in treatment axial disease in AS. However, they show some efficacy in management of peripheral and extra-articular features including psoriasis and uveitis. Non-steroidal anti-inflammatory drugs (NSAIDS) are the first line of medication for treatment of pain and stiffness in AS. NSAIDS such as naproxen provide some symptom relief and have been shown to significantly reduce inflammatory markers like c-reactive protein in blood of AS patients. However, long term NSAID treatment has the potential for gastrointestinal and cardiovascular toxicity.

The most effective symptom relief is obtained through use of the TNF-antagonist family of medications. Anti-TNF therapy provides substantial symptom relief and has effects on many of the axial, peripheral and extra-articular features of AS. However, while anti-TNF therapy reduces inflammation and prevents further joint destruction, it neither prevents nor slows bone formation. Side effects of long-term anti-TNF therapy include increased susceptibility to tuberculosis infection due to systemic immune suppression.

Since no treatment to date has been shown to have an effect on preventing or slowing the natural course of disease there is a very great need for research to determine what triggers disease and to better understand the inflammatory pathways involved in AS, thereby providing novel therapeutic options. Recent advance in the genetics and genomics of AS have highlighted several important pathways involved in AS. We review here some of those advances and highlight some of the functional consequences of the genetic associations with AS. Further genetic and gene expression studies can generate diagnostic algorithms to aid in early disease detection and improved treatment regimes.

**AS genetics**

As stated above, AS is a largely genetically controlled disease. Familial aggregation indicating the presence of shared susceptibility factors has been long observed, and studies of disease concordance in twins and families have shown that disease susceptibility is largely controlled by genetic factors. Twin studies have shown concordance to be 75% (6/8) in monozygotic twins, 27% (4/15) for dizygotic twins sharing HLA-B27 and 12.5% (4/32) for dizygotic twins overall. AS is highly heritable, with additive genetic factors estimated to account for >90% of susceptibility. Age of onset (40%) and disease severity, as measured by the widely accepted BASDAI (51%) and BASFI (76%) questionnaires, are also highly heritable. Below we discuss some of the more established loci contributing to this heritability.

**MHC**

The most significant AS-associated genetic locus is the major histocompatibility complex (MHC) on chromosome 6p. This genetic association is predominantly due to the association of AS with HLA-B27, however there is evidence for association with other MHC alleles.

**HLA-B27**

The association of HLA-B27 has been recognised since the early 1970s, and is among the strongest genetic...
associations with a common disease. HLA-B27 is carried by approximately 8% of white Europeans, compared with 3% to 5% of Han Chinese, and is rare in African populations (which accounts for the low incidence of AS in Africans). Around 80% to 95% of AS patients are HLA-B27 positive, yet only 1% to 5% of HLA-B27 carriers develop AS, therefore screening has not been useful as a sole diagnostic.

There are over 40 different allelic variants of HLA-B27 reported, all of which are ancestrally related to the B*2705 variant. 95% of white British HLA-B27 carriers have this subtype, while Chinese populations mainly have a mixture of B*2704 and B*2705. Of the many HLA-B27 subtypes identified, the majority are rare and their influence on AS has not been investigated.

Although the association between HLA-B27 and AS has been known for many years, the mechanism of disease association is still unclear. The "arthritogenic peptide" theory focuses on the canonical functions of HLA-B27 in antigen presentation to cytotoxic T-lymphocytes. The hypothesis suggests that after pathogenic insult, HLA-B27 presents a microbial epitope and elicits a CTL response against it. However, some of the CTLs may cross-react with any self antigens showing molecular mimicry with the microbial epitope. The reaction of CD8+ CTLs with self antigen would result in autoimmune tissue injury and inflammation. Association of ERAP1, the gene encoding the protein endoplasmic reticulum aminopeptidase 1 (ERAP1), with AS is only seen in HLA-B27 positive cases. This suggests that in HLA-B27- positive disease aberrant peptide trimming or presentation by ERAP1 and HLA-B27 are involved in AS pathogenesis, and this will be further discussed later.

Non-canonical theories involve properties of HLA-B27 itself. These include the "misfolded protein" hypothesis, in which the slow folding of the HLA-B27 heavy chain causes misfolded proteins that accumulate in the endoplasmic reticulum in the form of covalent homodimers and multimers. This accumulation of misfolded proteins causes ER stress, which results in inflammation through the activation of the unfolded protein response (UPR) and the ER-overload response (EOR).

An extension of the misfolding hypothesis is the HLA-B27 surface homodimer hypothesis. Covalently linked HLA-B27 homodimers have been found to be expressed on the cell surface, however they have not been secreted to the cell surface from the ER. Signalling through these non-canonical forms of HLA-B27 may play an immunomodulatory role and upset the normal development of HLA class-I responses, with unusual signalling causing excessive pro-inflammatory cytokine release. Additionally, HLA-B27 homodimer expressing APCs have been shown to stimulate the survival, proliferation and IL-17 production of Th17 cells that express the NK-receptor KIR3DL2 in AS patients. This observation provides a link between HLA-B27 and the IL-23 signalling pathway, which will be discussed later.

Other MHC genes
As well as the overwhelming association with HLA-B27, AS also shows association with other genes in the MHC. The MHC class I gene HLA-B60 has been shown to be associated with AS in a UK population. HLA-B60 shows association with both B27-positive and negative AS, and this association has been replicated in a Chinese population.

The MHC class II gene HLA-DRB1 has been found to be associated with AS in B27 matched case control studies and twin studies. A study of non-B27 MHC associations of AS using B27-matched haplotypes in cases and controls has shown that HLA-DRB1 is associated with AS irrespective of whether the haplotype contained HLA-B27. This may contribute to disease by causing aberrant presentation to or selection of CD4+ lymphocytes.

Non-MHC genes
Very little progress was made in the identification of genes associated with AS in the years following the discovery of the HLA-B27 association. A number of linkage studies were performed, identifying association with the interleukin-1 (IL-1) gene complex and CYP2D6. There have been several reports of IL-1 complex association with AS, including reported association in a Chinese population. The IL-1 gene complex includes genes encoding the pro-inflammatory cytokines IL-1α, IL-1β and their naturally occurring inhibitor IL-1 receptor agonist (IL-1RA). IL-1 is involved in the induction of chemokine and adhesion molecule expression in various cell types, and variation within the IL-1 cluster may contribute to AS pathogenesis through altered inflammation.

Association with AS has also been found for CYP2D6, the gene that encodes cytochrome P450 debrisoquine 4-hydroxylase. It is possible that CYP2D6 is involved in the metabolism of a ubiquitous environmental factor involved in AS, and that reduced enzymatic activity could lead to increased disease susceptibility, however this association has not been seen in more recent studies.

In recent years there have been a number of studies in which non-MHC genes have been reported as being associated with AS, most of which have been undertaken in populations
of white-European heritage with confirmed disease according to the Modified New York Criteria. The first study to convincingly identify non-MHC associations with AS was a study of 14,500 non-synonymous SNPs performed by the Wellcome Trust Case Control Consortium (WTCCC) and the Australo-Anglo-American Spondyloarthritis Consortium (TASC), in which associations with ERAP1 and IL23R were identified\textsuperscript{75} in a cohort of 1000 AS patients and 1500 healthy controls. The first genome-wide association study (GWAS) in AS was performed by TASC, using Illumina HumHap370 genotyping chips in a cohort of 2500 AS patients and 5000 healthy controls, with replication in a cohort of 900 cases and 1500 controls. This study identified the genes IL1R2 and ANTXR2 as being associated with AS, as well as gene deserts at chromosomes 14, 17, 21, 12, 9, 6, 5, 4, 2, and 1.

Abbreviation: GWAS, genome-wide association study.

### Confirmed Candidate Genes Associated with AS with Genomewide Significance Identified Through GWAS Studies\textsuperscript{1,3}

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP rsID</th>
<th>Putative gene of interest</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p31</td>
<td>rs11209026</td>
<td>IL23R</td>
<td>( 2.3 \times 10^{-17} )</td>
</tr>
<tr>
<td>1q32</td>
<td>rs2297909</td>
<td>KIF21B</td>
<td>( 5.2 \times 10^{-12} )</td>
</tr>
<tr>
<td>1p36</td>
<td>rs11249215</td>
<td>RUNX3</td>
<td>( 9.2 \times 10^{-11} )</td>
</tr>
<tr>
<td>2q11.2</td>
<td>rs2310173</td>
<td>IL1R2</td>
<td>( 4.8 \times 10^{-7} )</td>
</tr>
<tr>
<td>2p15</td>
<td>rs1085331</td>
<td>–</td>
<td>( 6.5 \times 10^{-24} )</td>
</tr>
<tr>
<td>4q21</td>
<td>rs4389526</td>
<td>ANTXR2</td>
<td>( 9.4 \times 10^{-8} )</td>
</tr>
<tr>
<td>5q15</td>
<td>rs30187</td>
<td>ERAP1</td>
<td>( 1.8 \times 10^{-27} )</td>
</tr>
<tr>
<td>5q33</td>
<td>rs6556416</td>
<td>IL12B</td>
<td>( 1.9 \times 10^{-8} )</td>
</tr>
<tr>
<td>5p13</td>
<td>rs10440635</td>
<td>PTGER4</td>
<td>( 2.6 \times 10^{-7} )</td>
</tr>
<tr>
<td>6p21</td>
<td>rs449859</td>
<td>HLA-B</td>
<td>( &lt;10^{-200} )</td>
</tr>
<tr>
<td>9q34</td>
<td>rs10781500</td>
<td>CARD9</td>
<td>( 1.1 \times 10^{-6} )</td>
</tr>
<tr>
<td>12p13</td>
<td>rs11616188</td>
<td>TNFRSF1A</td>
<td>( 4.1 \times 10^{-12} )</td>
</tr>
<tr>
<td>17q21</td>
<td>rs8070463</td>
<td>TBKBP1</td>
<td>( 5.3 \times 10^{-8} )</td>
</tr>
<tr>
<td>21q22</td>
<td>rs378108</td>
<td>–</td>
<td>( 2.1 \times 10^{-11} )</td>
</tr>
</tbody>
</table>

**Table 1** Confirmed candidate genes associated with AS with genomewide significance identified through GWAS studies.

ERAP1 was first reported to be associated with AS in the WTCCC/TASC nsSNP study,\textsuperscript{76} and this association has been confirmed both in the TASC\textsuperscript{76} and WTCCC2/TASC GWAS,\textsuperscript{37} as well as in a number of different ethnicities including Chinese,\textsuperscript{77,78} Portuguese,\textsuperscript{90} Korean,\textsuperscript{81} and Hungarian.\textsuperscript{37} ERAP1 has also been shown to be associated with psoriasis in both white European\textsuperscript{82} and Chinese\textsuperscript{83} populations.

There are two known roles of ERAP1, both of which may explain the association with AS. ERAP1 is involved in the trimming of peptides within the endoplasmic reticulum to optimal length for MHC class I presentation.\textsuperscript{84,85} The strong association with HLA-B27 indicates that AS is primarily an MHC class I mediated disease, but the way B27 increases disease susceptibility is currently unknown. If the association of ERAP1 with AS relates to peptide presentation, this may help explain the association with HLA-B27 (Figure 3). The second function of ERAP1 is the cleavage of cell surface receptors. ERAP1 has been shown to cleave the cell surface receptors for the pro-inflammatory cytokines IL-1 (IL-1R\textsuperscript{2}), IL-6 (IL-6R\textsuperscript{3}), and TNF (TNFRSF1A),\textsuperscript{88} downregulating their signalling. Interestingly, both IL1R2\textsuperscript{75} and TNFRSF1A\textsuperscript{77} have been shown to be associated with AS. Genetic variants altering the receptor shedding function of ERAP1 could potentially have pro- or anti-inflammatory effects through this mechanism, however studies in mice have shown no difference in receptor shedding between *Erap1*\textsuperscript{75} and C57BL/6 control mice.\textsuperscript{37}

The primary AS-associated SNP in ERAP1 is rs30187, which is an nsSNP encoding the protective p.Lys528Arg amino acid change. There is also an independent secondary association observed with the nsSNPs rs17482078 (p.Arg725Gln) and rs10050860 (p.Asp575 Asn), however these two variants are in complete linkage disequilibrium (LD) and cannot be split genetically.\textsuperscript{57} The crystal structure of *ERAP1* has recently been published,\textsuperscript{89} and has revealed the location of these altered residues around the substrate binding and regulatory sites of ERAP1. Peptide trimming assays performed using recombinant ERAP1 in vitro have shown that the protective rs30187 and rs17482078 variants had ~40% slower rates of trimming than wild-type,\textsuperscript{57} suggesting that they are both loss-of-function variants. It has also been shown that while wild-type ERAP1 is able to trim peptides to 8- or 9-mers, the rs30187 variant stops at 11-mers.\textsuperscript{90}

Gene-gene interaction between HLA-B27 and ERAP1 has been observed, with *ERAP1* association only observed with AS in HLA-B27 positive individuals.\textsuperscript{57} This restriction in association supports the models in which abnormal peptide trimming or presentation by *ERAP1* and HLA-B27 contribute to HLA-B27 disease. This suggests that inhibition of ERAP1 may be effective in HLA-B27 positive AS.
RUNX3 is a key gene involved in CD8 lymphocyte differentiation. This gene was identified as being associated with AS in the TASC/WTCCC2 GWAS, and this association supports the hypothesis that HLA-B27 contributes to AS pathogenesis through a mechanism involving peptide presentation to CD8 lymphocytes. It has been shown that AS cases have reduced CD8 lymphocyte counts, as do healthy individuals who carry the AS-associated SNPs. These findings suggest that CD8 lymphocytes are involved in AS pathogenesis, likely in response to presentation of antigenic peptides by HLA-B27 after antigen processing by ERAP1.

**IL-23 signalling pathway genes – IL23R, IL12B, STAT3**

IL23R encodes the receptor for the cytokine interleukin 23 (IL-23), which is a key regulatory cytokine essential for the differentiation of Th17 cells (Figure 3). The association of IL23R suggests that Th17 cells are involved in the pathogenesis of AS in white European populations, as well as other IL23R associated diseases. IL23R has been found to be associated with a number of diseases in White European populations, including Crohn’s disease, psoriasis, and psoriatic arthritis. These diseases are all closely clinically related members of the seronegative group of autoimmune diseases, and this suggests that IL23R may partially explain their coexistence.

The association of IL23R was first reported in the 14,500 nsSNP study and has been subsequently confirmed by two GWAS, as well as in Portuguese populations. Studies in Han Chinese and Koreans have shown no association with IL23R, and in these studies the primary associated nsSNP in white Europeans, rs11209026, was observed to be non-polymorphic. However, it is possible that there are rare AS-associated variants of IL23R in East Asian populations, and we have next-generation sequencing data that suggests that this is the case in a Han Chinese population (submitted for publication).

The primary AS-associated nsSNP in IL23R codes for a Gln381Arg substitution in exon 9, and this variant is protective against disease. There is also an independent secondary association with rs11209032, which lies upstream of IL23R, and this variant increases disease susceptibility. There have
been studies in which the function of the p.Gln381 Arg variant have been investigated.98,99 In these studies it has been shown that the p.Gln381Arg variant was a loss-of-function change and resulted in decreased levels of IL-23 signalling. This resulted in lower levels of IL-23 mediated IL-17 and IL-22 production, as well as lower levels of circulating Th17 and Tc17 cells. This is consistent with the increased Th17 lymphocyte numbers55,100,101 and serum IL-17 levels102,103 observed in AS patients. However, IL23R is also expressed on a number of other cell types, including γδ T cells, NK cells and NKT cells104–106 and it is not clear which cell type is primarily affected by the disease-associated variant. Our data has shown there are increased levels of IL-23R+ γδ T cells in the peripheral blood of AS patients compared to healthy controls, suggesting that these cells may play a role in disease pathogenesis.17

The IL-12p40 subunit shared between IL-12 and IL-23 is encoded by the gene IL12B. This gene was previously reported as having suggestive association with AS,7 and this association was later confirmed at the genome-wide level,57 further implicating the role of IL-23 signalling in AS pathogenesis. IL12B has also been previously reported as being associated with both Crohn’s disease and psoriasis.91,93,107

STAT3 has been shown to be associated with AS in both white European7 and Han Chinese78 populations, as well as being previously shown to be associated with Crohn’s disease.108 STAT3 is a critical transcription factor in the differentiation of Th17 cell populations, and is found directly downstream of IL-23R in the IL-23 signalling cascade (Figure 3).109,110 Loss of function mutations in STAT3 cause Job syndrome (OMIM 147060), which is characterised by a loss of Th17 cells and is associated with recurrent, severe infections with extracellular bacteria and fungi.109–111

Fungal response – CARD9 and PTGER4
CARD9 and PTGER4 were both identified as being associated with AS in the TASC/WTCCC2 GWAS57 and have both previously been shown to be associated with Crohn’s disease.112,113 CARD9 mediates signals from the innate immunity receptors dectin-1 and -2, which recognise β-glucan, a component of fungal and some bacterial cell walls. This induces the production of PGE2, which is the ligand for the prostaglandin E receptor 4, EP4 subtype, which is encoded by PTGER4. PTGER4 then induces the production of IL-23 and IL-17, promoting the expansion of Th17 lymphocytes (Figure 3).114 It has been reported that when SKG mice are treated with β-glucan they develop spondyloarthritis and Crohn’s disease.115 and that this model is characterised by activation of Th17 cells.116 This model suggests that involvement of the IL-23 signalling pathway in AS pathogenesis may be triggered by ubiquitous pathogens carrying β-glucan. PTGER4 has also been shown to play an anabolic role in bone remodelling,117,118 making it a good candidate for a role in the bone formation that characterises AS.

Other associated genes
IL1R2
The IL-1 gene cluster on chromosome 2q has been repeatedly reported as being associated with AS, with association observed with IL1A.119 Neither the WTCCC/TASC nsSNP study nor the TASC or TASC/WTCCC2 GWAS identified association at the IL-1 gene cluster, however the TASC GWAS did identify and confirm association with IL-1 receptor, type II, IL1R2.76 Although the strongest association was observed with IL1R2, the association peak is quite broad, and may also include association at IL1R1.

IL-1R2 binds IL-1α and IL-1β with high affinity, acting as a decoy receptor after cleavage from myeloid and lymphoid cell membranes, possibly by ERAP1.86 Further investigation is required to determine the disease mechanism.

ANTXR2
ANTXR2 is the gene encoding the protein capillary morphogenesis protein-2 (CMG2), and was identified as being associated with AS in both the TASC76 and the TASC/WTCCC2 GWAS.95 Recessive mutations of ANTXR2 cause juvenile hyaline fibromatosis (OMIM 228600) and infantile systemic hyalinosis (OMIM 236490), but it has not been previously associated with any complex diseases. It is not clear how this gene functionally contributes to AS pathogenesis.

2p15 and 21q22 gene deserts
The 2p15 and 21q22 intergenic regions or “gene deserts” have been found to be strongly associated with AS in both the TASC76 and TASC/WTCCC2 GWAS. There are no genes within 100 kb of the 2p15 peak associated region. The proteasome assembly chaperone 1 gene PSMG1 lies nearby the 21q22 locus, however it is separated by a recombination hotspot and it is therefore unlikely that any SNP at 21q22 is in LD with any SNP in PSMG1. Possible mechanisms by which these associations contribute to disease include long-range cis-acting regulatory regions controlling expression of nearby genes, or effects on non-coding RNA.
**KIF21B**

The gene **KIF21B** encodes for a member of the family of kinesin motor proteins, and has been previously shown to be associated with Crohn’s disease. **KIF21B** was first identified as being associated with AS in a study in which genes previously shown to be associated with Crohn’s disease were investigated in an AS cohort, and was confirmed in the TASC/WTCCC2 GWAS. Kinesins are used for essential component transport along dendritic and axonal microtubules by neurons. The Kinesin family member **KIF5A** has been associated with rheumatoid arthritis and type-1-diabetes, as well as being close to a locus associated with multiple sclerosis. If the association of **KIF5A** is confirmed as true, this suggests alternate roles for kinesin family members.

**TNFRSF1A**

The association of **TNFRSF1A** further supports the involvement of TNF pathways in ankylosing spondylitis pathogenesis. In addition to showing association with AS, **TNFRSF1A** has been previously shown to be associated with both Crohn’s disease and ulcerative colitis. Additionally, a mouse model with constitutive over expression of TNF leads to inflammatory bowel disease, and spondyloarthropies resembling AS, in a manner dependent on the expression of TNFRI, supporting a role for **TNFRSF1A** in AS pathogenesis. **TNFRSF1A** has also been shown to be associated with AS in a Han Chinese population, increasing the likelihood of **TNFRSF1A** being a true AS susceptibility gene.

Association has also been observed with **TBKBP1**, which encodes the TNF receptor signalling pathway member TBK binding protein. Suggestive association has also been observed with **TRADD**, the TNF receptor associated death domain protein, and further research is needed to elucidate the actual disease mechanism responsible for these associations.

**Genomic studies**

**Expression profiling**

Although significant effort has gone into genetic studies of the spondylarthropies, particularly AS including the several large scale GWAS studies described above, large scale gene expression profiling studies have been somewhat less extensive. Other similarly prevalent inflammatory conditions such as RA and Lupus have been the subject of much more extensive study, with over 10 such studies in both diseases in the last two years alone.

Undertaking whole genome expression studies generates powerful datasets that can inform on a number of disease aspects;

1. Identification of genes involved in the disease pathology which can inform on;
   a. The genes directly contributing to disease
   b. Identify pathways that can be targeted therapeutically
   c. Identify environmental factors that contribute to pathology
2. Characterise gene changes through disease progression to identify specific gene involvement at different disease stages
3. Elucidate cell- or tissue-specific genes participating in the disease process
4. Identify biomarkers which can be measured to:
   a. Diagnose early disease
   b. Catalog disease progression
   c. Monitor treatment response.

**Expression studies in ankylosing spondylitis**

In most large-scale genetic studies germline DNA is used which can be simply extracted from blood or saliva samples. For the majority of these studies in relatively common diseases it is assumed causative mutations are germline making the tissue source less of a consideration. In cases of rarer genetically mosaic conditions such as Melorheostosis, issue of tissue source become extremely important. For expression profiling studies however sample source and handling are always key considerations. The fact that expression profiling “snapshots” the cellular activity in the sample requires consideration to be given to the;

1. Disease and treatment status of the individual
2. Relevance of sample tissue site to disease presentation
3. Handling of tissue sample to minimise potential RNA degradation.

As described earlier there are two aspects to AS pathology, an initial systemic inflammation and the osteoproliferation and joint ankylosis characterising late disease. To monitor these quite different processes, careful consideration has to be given to the suitability of tissue sample. Systemic inflammation can be studied looking at the immune component of the blood, either through RNA extracted from whole blood or the peripheral blood mononuclear cell (PBMC) fraction which are relatively accessible. To study the joint disease processes is considerably more difficult. Biopsies of inflamed or ankylosed joints are rarely performed.
In cases of severe joint fusion, tissue may be obtained during surgery and usually only warranted in cases of very advanced disease where early key disease-causing changes will not be present.

**Disease-site studies**

Due to the rarity of informative axial disease-site samples in AS only one expression profiling study in these tissues has been reported. Using sacroiliac joint aspirates from SpA patients, Rihl et al identified sacroiliitis candidate genes and then validated these candidates in knee synovial fluid biopsies from patients with peripheral synovitis.\(^{126}\) They identified *interleukin-7* (*IL7*) as the best candidate gene but the small size of the study, only 3 spondyloarthropy (SpA) cases and no axial tissue controls, and the small gene array (1200 genes) means these preliminary findings require further confirmation.

Despite axial inflammation being the hallmark of AS, significant peripheral pathology is also found in many patients. Knee synovial biopsies present a more accessible and less invasive tissue candidate for assessment of expression profiles at joint disease sites. To date only two small scale studies have been published using “macroarrays” (600 or 1200 genes). Both of these studies emanated from David Yu’s group at UCLA, one comparing osteoarthritis (OA) and SpA knee synovial biopsies\(^{127}\) and the other comparing SpA and RA synovial fluid mononuclear cells (SFMCs).\(^{128}\) In the SFMC study, similar patterns of inflammatory cytokines were expressed. One strong candidate gene was identified, the unfolded protein response (UPR) gene, *immunoglobulin heavy chain binding protein* (*BiP*). The UPR has been proposed as a mechanism by which *HLA-B27* might mediate its role in AS.\(^{129}\) In the synovial biopsy study, SpA and OA samples could be differentiated by their expression profiles, however candidates were not validated.

Our group has recently undertaken an expression profiling study in SpA synovial biopsies using a different approach. Rather than using fresh-frozen biopsies we have utilised Illumina’s Whole-Genome DASL\(^{8}\) (cDNA-mediated Annealing, Selection, Extension, and Ligation [Illumina Inc, San Diego, CA]) Gene Expression Assay which has been specifically developed for whole-genome expression profiling of degraded RNA samples from archived tissue biopsies. We analyzed the expression profile of 24000 cDNAs in synovial biopsy samples from seronegative spondyloarthritis, AS and osteoarthritis patients and normal control samples (submitted for publication). Several interesting candidate genes were identified. *Matrix metalloproteinase 3* (*MMP3*), a gene associated with inflammation and bone and cartilage remodelling was strongly upregulated and immunohistochemical confirmation showed MMP3 expression was particularly high in AS biopsies. *MMP3* has previously been identified as a circulating biomarker of AS.\(^{23,130,131}\) *Dickkopf-3* (*DKK3*), a Wnt pathway inhibitor, was downregulated. With the Wnt pathway having been recently hypothesised as playing a key role in AS-associated osteoproliferation\(^{132–134}\) this was particularly interesting. Gene ontology analysis also showed altered immune and inflammation pathways and more interestingly changes in extracellular matrix and osteoblast associated pathways.

Use of peripheral joint samples for expression profiling in AS is not ideal and interpretation of these findings must be undertaken carefully. Unfortunately however axial samples are extremely rare. They are certainly valuable in identifying unique profiles for AS in relation to RA and/or OA samples to eliminate general inflammatory/joint damage genes. Further, many of the early changes in inflammation-mediated cell activity initiating tissue damage would be similar in both peripheral and axial joints.

With the limited number of disease-site studies having been undertaken in AS/SpA, little consensus on key genes have been found. Even using peripheral joint samples, experimental cohorts are small and usually lack sufficient numbers to allow independent validation. To generate sufficient power, consortium approaches should be considered and utilisation of technologies such as DASL to probe archived samples would further increase the numbers of samples available to profile.

**Circulating cell studies**

As described above, there is also a systemic immune-mediated inflammatory component in AS in addition to the osteoproliferation seen in the axial joints. Thus peripheral blood presents an inviting target tissue to try and elucidate the underlying gene changes that contribute to the early aspects of AS/SpA. The comparative ease with which samples of peripheral blood can be obtained has also meant that expression profiling studies using this RNA source in AS/SpA (as well as many other immune-mediated diseases) have been better powered and more numerous than those described above in joint tissue.

In AS the circulating immune cells may reflect the early changes that contribute both to disease susceptibility and also early stage progression. Most studies to date have utilised the peripheral blood mononuclear cell (PBMC) fraction from
whole blood which contains the circulating white blood cells including B, T and NK cells, monocytes and granulocytes. Although a widely utilised standard protocol, purification of PBMCs does take several hours and needs to be carried out within 48 hours of blood collection, preferably within a few hours. This is particularly true for transcriptomics where differences in cell handling could cause significant shifts in gene expression patterns. PBMCs can therefore by unsuitable for studies involving multiple centers or tissue sampling from more remote or distant locations. An alternative blood-derived RNA source is to use whole blood. Two kits have been specifically developed for transcriptomic studies in whole blood with the blood collected into an RNA preservative which stabilises the RNA allowing shipping/initial storage at room temperature with minimal degradation (PAXgene from Qiagen [Valencia, CA] or Tempus from Invitrogen/Applied Biosystems [Life Technologies, Grand Island, NY]).

Three whole genome expression profiling studies have been published using whole blood in spondylarthropy all using the PAXgene technology. Sharma et al compared healthy controls and spondyloarthropy patients but used a slightly unusual approach. Rather than analysing a main sample group by microarray then confirming candidate genes by qPCR in a second cohort, they split the sample groups into a discovery and a validation set and micro arrayed both groups and compared differentially expressed genesets. However, the control and patient group had significant imbalances in their gender mix, as well as no joint disease scores being ascertained making interpretation difficult. Both Assassi et al and Pimental-Santos et al focussed on AS, using patients with confirmed AS according to the Modified New York Criteria, with Assassi also comparing their datasets with those from systemic lupus erythematosus (SLE) and systemic sclerosis (SSc) patients. Both these studies undertook an initial microarray study and confirmed candidate genes by qPCR in a larger 2nd cohort. The Pimental-Santos study was the largest study with 18 patients and 18 controls in the initial microarray cohort and validated in 156 patients and controls using Taqman Low Density Arrays (TLDAs). Interestingly all three studies cohort and validated in 156 patients and controls using Taqman study with 18 patients and 18 controls in the initial microarray a larger 2nd cohort. The Pimentel-Santos study was the largest microarray study and confirmed candidate genes by qPCR in a second cohort, they split the sample groups into this gene maybe generally associated with inflammatory processes, was confirmed.

Only recently however have large-scale whole genome expression profiling studies on the complete PBMC population been undertaken in AS or SpA by ourselves and David Yu’s group at UCLA. Using Illumina’s HT-12 whole genome arrays we compared expression profiles of PBMCs from 18 active AS patients (confirmed by Modified New York Criteria) and 18 age- and gender-matched controls identifying a 452-gene signature delineating AS and control samples with 94% accuracy. Downregulation of four immune/inflammation candidate genes, NR4A2, TNFAIP3, CD69, and RORA, was validated by qPCR. The expression pattern of these four genes generated a diagnostic potential of ~80% for AS which has the potential to be used to distinguish inflammatory from non-inflammatory back pain.

The downregulation of these candidate genes, together with other immune genes, suggests a defect in the capacity of the immune response. In most inflammatory conditions an upregulation of inflammation-associated genes would be expected, such as in RA and systemic lupus erythematosus. Our results however are consistent with the macrophage study described above which also detailed a decreased immune response. They also saw a decrease in TNFAIP3 expression in AS patients.
Down-regulation or muting of the immune response is compatible with one of the hypotheses proposed for the underlying pathology causing AS, in which a ubiquitous environmental factor acts as a trigger in patients genetically disposed to the disease (such as carrying the HLA-B27 allele together with other susceptibility alleles from other genes such as IL23R and ERAP1). A microarray study on whole blood samples from psoriatic arthritis patients, a related spondyloarthropathy, also showed a general down-regulation in immune-associated genes suggesting inappropriate immune control.

In addition to comparing AS and control groups, Gu et al also looked at undifferentiated SpA (uSpA) patients. Expression changes were significantly greater in the uSpA patients compared to controls than for the AS samples. Given that uSpA is considered a less severe disease this was somewhat unexpected. Very few genes altered in AS with most of those being up-regulated including NR4A2 which was downregulated in our study.

The Gu et al study was undertaken in a Chinese patient cohort which might explain the different findings to our Caucasian cohort study.

Table 2 summarises the most interesting genes identified through whole genome expression profiling studies in well-powered studies.

As mentioned above, expression profiling can generate powerful diagnostic algorithms. In AS, advanced joint damage is required for firm diagnosis (Modified New York Criteria) to gain subsidized access to the most effective (and expensive) treatments. A need for a reliable early diagnostic in AS is thus very pressing. Both ours and David Yu's studies showed diagnostic power > 80% to predict disease based on the expression profiles.

The results to date have shown a disappointing lack of overlap between the whole blood and PBMC expression studies. Several studies have reported differences in the expression profile of PBMCs vs whole blood. Larger cohort studies are required with whole blood profiling the more convenient vehicle to expedite the multicenter approach needed to generate sufficient power.

To date a number of disparate expression profiling studies have been undertaken in AS. Although a number of potentially interesting gene changes have been identified, there is a lack of consensus in the mechanisms highlighted save for an indication of a reduced immune response. A similar problem confronted genetic studies several years ago and this was overcome by large-scale collaborative studies generating large homogenous sample cohorts and enabled strong reproducible candidate gene-associations to be identified. Similar issues need to be addressed for future expression profiling studies in AS. Significant thought must be given to assembling large informative sample cohorts that reflect the stage of disease being investigated. These cohorts need to consist of homogenous case collections targeting specific disease

### Table 2 Candidate genes associated with AS identified through genome-wide expression studies

<table>
<thead>
<tr>
<th>Gene</th>
<th>Samples</th>
<th>Tissue</th>
<th>Platform</th>
<th>Regulation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMP3</td>
<td>AS-SpA vs normal-OA</td>
<td>Synovial biopsy</td>
<td>Illumina DASL</td>
<td>Up</td>
<td>Submitted for publication</td>
</tr>
<tr>
<td>DKK3</td>
<td>AS-SpA vs normal-OA</td>
<td>Synovial biopsy</td>
<td>Illumina DASL</td>
<td>Down</td>
<td>Submitted for publication</td>
</tr>
<tr>
<td>TLR4</td>
<td>AS vs controls</td>
<td>Whole blood (PAXGene)</td>
<td>Illumina HT12</td>
<td>Up</td>
<td>135,136</td>
</tr>
<tr>
<td>TLR5</td>
<td>AS vs controls</td>
<td>Whole blood (PAXGene)</td>
<td>Illumina HT12</td>
<td>Up</td>
<td>135,136</td>
</tr>
<tr>
<td>NLRP2</td>
<td>SpA vs controls</td>
<td>Whole blood (PAXGene)</td>
<td>Affymetrix HG-U133 plus 2.0</td>
<td>Down</td>
<td>137</td>
</tr>
<tr>
<td>CLEC4D</td>
<td>AS vs controls</td>
<td>Whole blood (PAXGene)</td>
<td>Illumina HT12</td>
<td>Up</td>
<td>136</td>
</tr>
<tr>
<td>SPARC</td>
<td>SpA vs controls AS vs controls</td>
<td>Whole blood (PAXGene)</td>
<td>Affymetrix HG-U133 plus 2.0</td>
<td>Up</td>
<td>136,137</td>
</tr>
<tr>
<td>IFN-regulated genes</td>
<td>AS vs controls</td>
<td>Purified macrophages</td>
<td>Affymetrix HG-U133 plus 2.0</td>
<td>Reverse</td>
<td>IFN signature 142</td>
</tr>
<tr>
<td>LIGHT</td>
<td>AS patients with and without anti-TNF treatment</td>
<td>Whole blood</td>
<td>Affymetrix HG-U133 plus 2.0</td>
<td>Down</td>
<td>143</td>
</tr>
<tr>
<td>NR4A2</td>
<td>AS vs controls</td>
<td>PBMC</td>
<td>Illumina HT12</td>
<td>Down</td>
<td>145</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>AS vs controls</td>
<td>PBMC</td>
<td>Illumina HT12</td>
<td>Down</td>
<td>142,145</td>
</tr>
<tr>
<td>CD49</td>
<td>AS vs controls</td>
<td>PBMC</td>
<td>Illumina HT12</td>
<td>Down</td>
<td>145</td>
</tr>
<tr>
<td>RORA</td>
<td>AS vs controls</td>
<td>PBMC</td>
<td>Illumina HT12</td>
<td>Down</td>
<td>145</td>
</tr>
<tr>
<td>NR4A2</td>
<td>AS vs uSpA vs controls</td>
<td>PBMC</td>
<td>Illumina Ref-8</td>
<td>Up</td>
<td>146</td>
</tr>
<tr>
<td>RGS1</td>
<td>AS vs uSpA vs controls</td>
<td>PBMC</td>
<td>Illumina Ref-8</td>
<td>Up</td>
<td>146</td>
</tr>
</tbody>
</table>

**Abbreviations:** PBMC, peripheral blood mononuclear cells; AS, ankylosing spondylitis; SpA, spondyloarthritis; uSpA, undifferentiated spondyloarthritis; OA, osteoarthritis; IFN, interferon.
stages, such as early inflammation or late-stage ankylosis, to stand the best chance of being informative. For the development of diagnostic tests, further studies need to be conducted on the suitability of whole blood RNA as a sample source reflective of disease status, as this approach is much more practical for use in clinical practice. Given the scarcity of clinically relevant samples, collaborative approaches would enable such powerful studies. Archived samples should also be utilized using the appropriate technologies developed for such approaches.

With next generation sequencing enabling whole transcriptome studies the future of expression profiling looks to lie in this direction. Costs are rapidly decreasing making large scale studies viable for well-funded labs. RNAseq will allow the full catalog of splice variants, non-coding RNAs and microRNAs to be elucidated and some of these novel species may explain some of the strong GWAS hits in “gene deserts”.

**Conclusion**

Although a number of strong candidate genes have been identified from well-powered GWAS studies there appears to be little correlation with the gene expression profiling studies published to date. This could be due to;

1. Genetic studies identify for the most part susceptibility genes and expression studies are frequently in patients with established disease. Many alterations in these genes may be finished by the time the expression studies are undertaken
2. Genetic studies may identify genes in which function rather than absolute expression level is altered
3. Genetic disposition to a disease may be the result of very small changes in a number of genes, each of which is below the threshold of detection by expression profiling but cumulatively might generate significant downstream effects which are picked up by the expression studies.

To date no expression quantitative trait loci (eQTL) studies have been published in AS or SpA. There is now a significant body of GWAS and expression profiling data to enable such studies. Such approaches correlating genetic variants with both *cis-* and *trans* expression may further explain to current GWAS/expression profiling disconnect and further reveal the pathways and mechanisms underlying AS.

**Disclosure**

None of the authors report any conflicts of interest.

**References**

Advances in Genomics and Genetics downloaded from https://www.dovepress.com/ by 54.70.40.11 on 07-Dec-2018


67. Wei JCC, Tsai WC, Lin HS, Tsai CY, Chou CT. HLA-B60 and B61 are strongly associated with ankylosing spondylitis in HLA-B27-negative Taiwanese patients. *Rheumatology.* July 1, 2004;43(7):839–842.


