

## ORIGINAL ARTICLE

# Epigenetic and gene expression analysis of ankylosing spondylitis-associated loci implicate immune cells and the gut in the disease pathogenesis

Z Li<sup>1</sup>, K Haynes<sup>2</sup>, DJ Pennisi<sup>1</sup>, LK Anderson<sup>1</sup>, X Song<sup>1</sup>, GP Thomas<sup>3</sup>, T Kenna<sup>1</sup>, P Leo<sup>1</sup> and MA Brown<sup>1</sup>

Ankylosing spondylitis (AS) is a common immune-mediated arthropathy primarily affecting the spine and pelvis. Most AS patients have subclinical intestinal inflammation, suggesting the gut microbiome and the immune response play a role in pathogenesis. Susceptibility to AS is primarily genetic, and at least 114 susceptibility variants have been identified to date. We applied bioinformatic methods utilizing epigenetic and gene and protein expression data to identify the cell types through which AS-associated variants operate. Variants were enriched in transcriptionally regulated regions in monocytes, CD4<sup>+</sup> and CD8<sup>+</sup> T cells, natural killer cells, regulatory T cells and B cells and mucosa from the small intestine, sigmoid colon and rectum. Weak signals were detected in bone cells, consistent with bone disease being a secondary manifestation. RNA sequencing of blood cells from AS patients and controls identified differentially expressed genes. Interrogation of expression databases showed that the upregulated genes were enriched in monocytes and downregulated genes were enriched in CD8<sup>+</sup> T cells and natural killer cells. Gene Ontology term enrichment analysis identified microbes and the gut in the aetiology of AS. These findings identify the key immune cell types that drive the disease, and further highlight the involvement of the gut microbiome in the pathogenesis of AS.

*Genes and Immunity* (2017) 18, 135–143; doi:10.1038/gene.2017.11; published online 15 June 2017

## INTRODUCTION

Ankylosing spondylitis (AS) is a common, chronic inflammatory disease that not only primarily affects the spine and the sacroiliac joints, but can also affect peripheral joints and extraarticular tissues including the eye, gut and skin. Furthermore, up to 70% of AS patients have subclinical intestinal inflammation, with 5–10% developing clinically diagnosed inflammatory bowel disease (IBD).<sup>1</sup> AS is highly familial and heritable,<sup>2,3</sup> with a prevalence of 0.55% in European populations<sup>4</sup> and 0.2–0.54% in Han Chinese populations.<sup>5</sup> Although the disease is strongly associated with *HLA-B27* within the major histocompatibility complex (MHC; odds ratio > 50),<sup>6,7</sup> *HLA-B27* contributes only ~20% of the genetic risk.<sup>8</sup> At least 39 genetic variants lying outside the MHC have been identified as associated with AS with genome-wide levels of significance.<sup>8–12</sup> A further 74 non-MHC associations have been identified through studies of pleiotropic genetic associations of AS, inflammatory bowel disease, psoriasis and primary sclerosing cholangitis.<sup>8</sup> Conditional analyses have revealed multiple independent human leukocyte antigen (HLA) associations with AS in addition to *HLA-B27*; for the purposes of this paper we consider this locus as one.<sup>13</sup>

As with most common variant disease and trait-associated polymorphisms, most of the associated single-nucleotide polymorphisms (SNPs) lie within noncoding regions.<sup>14,15</sup> Utilizing the increasing depth of genomic annotation of epigenetic and gene expression properties of different cell types, several recent studies have shown that disease- or trait-associated SNPs often operate through effects on gene regulatory regions in specific cell types or

tissues.<sup>16,17</sup> With the rapid increase in epigenetic and gene expression data in hundreds of cell types, associated variants can be further interrogated, and potential causal variants and loci prioritized for further study, focussing on the cell types in which they primarily operate.<sup>18–20</sup>

A wide variety of cell types have been suggested to be involved in AS, largely through flow cytometric analyses of cell counts and activation, and include CD4<sup>+</sup> and CD8<sup>+</sup> T lymphocytes;<sup>21</sup> T helper 17 (Th17) cells;<sup>22</sup> regulatory T cells;<sup>21</sup> innate immune cells, including natural killer (NK) cells,<sup>23</sup> gamma-delta cells and innate lymphoid cells;<sup>24</sup> B cells;<sup>25</sup> mast cells;<sup>11,26</sup> and intestinal Paneth cells.<sup>27–29</sup> Although these studies provide valuable information about the cellular mechanisms involved in AS pathogenesis, they only indirectly inform how AS-associated SNPs affect the pathogenic mechanism of the disease. A further limitation of these studies is that they focussed on cell types that are easy to sample, such as circulating immune cells, and thus the involvement of other rarer cell types, particularly in sites not easily sampled, remain underinvestigated. To date, there have been very few studies that utilized AS-associated SNPs and multiomics data to identify the cell types involved in the pathogenesis of the disease. One that broadly examined autoimmune and inflammatory diseases used PICS (Probabilistic Identification of Causal SNPs) and the chromatin maps of different cell types identified T<sub>naive</sub>, various T helper cells, and colonic and duodenal mucosa as cell types likely to be involved in AS.<sup>30</sup> Another, using a smaller number of curated AS-associated loci, further implicated Th17 cells in AS pathogenesis.<sup>31</sup> Recently, a cross-disease study of a

<sup>1</sup>Translational Genomics Group, Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology at Translational Research Institute, Brisbane, Queensland, Australia; <sup>2</sup>University of Queensland Diamantina Institute, The University of Queensland, Brisbane, Queensland, Australia and <sup>3</sup>Research Office, Charles Sturt University, Wagga Wagga, New South Wales, Australia. Correspondence: Professor MA Brown, Translational Genomics Group, Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology at Translational Research Institute, Brisbane, Queensland 4102, Australia. E-mail: Matt.brown@qut.edu.au

Received 6 February 2017; revised 25 April 2017; accepted 27 April 2017; published online 15 June 2017

larger cohort identified additional genome-wide significant (GWS) AS-associated loci.<sup>8</sup> Here, we leveraged the expanded AS-associated SNP set from the cross-disease study, publicly available cell type-specific epigenetic, and gene and protein expression data to elucidate the critical cell types involved in AS pathogenesis.

## RESULTS

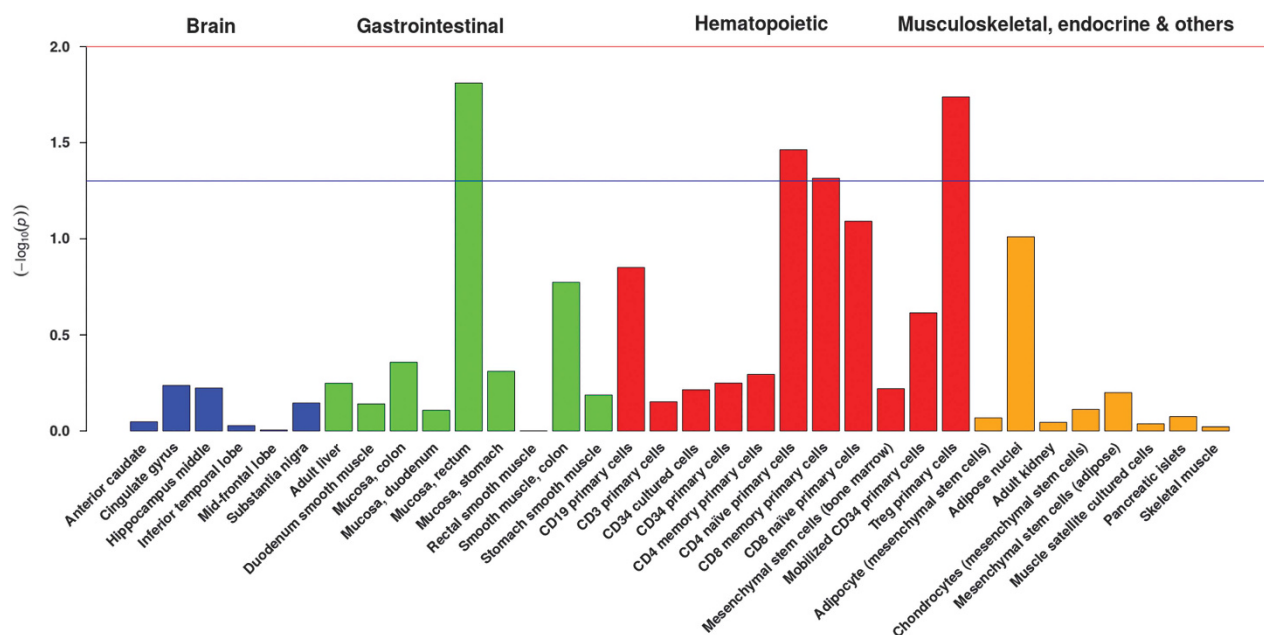
AS-associated SNPs are enriched in regulatory elements in gut and immune cell types

AS-associated SNPs were defined as either being GWS significant in analyses of AS alone (AS\_GWS\_40), or in cross-disease studies (AS\_cross\_disease\_114; see Materials and methods and Supplementary Information). Of the 39 most strongly AS-associated GWS (non-MHC) SNPs, only one was located in an exon, and of the 113 (non-MHC) SNPs from cross-disease studies, up to 13 were located in exons. There are 11 SNPs in linkage disequilibrium (LD;  $r^2 > 0.8$ ) with the 39 AS-GWS SNPs (28%), and 40 SNPs in LD with the 113 SNPs located in exonic region (35%). As 65–72% of disease-causative SNPs at these loci are likely to lie in noncoding regions, we examined whether any of these SNPs were located in regulatory elements of the genome as indicated by epigenetic marks. We began by assessing the enrichment of the AS-associated SNP data sets in H3K4me3 marks within 34 cell types using EPIGWAS.<sup>32</sup> AS GWS-associated SNPs showed significant enrichment in immune cell types and gut tissue, consistent with AS being an immune-mediated arthropathy in which the gut is involved in pathogenesis (Figure 1 and Supplementary Table S1). The highest levels of enrichment were observed in rectal mucosa ( $P$ -value = 0.0155) and regulatory T cells ( $P$ -value = 0.0183). CD4-naïve T cells and CD8 memory T cells also showed signals. With the larger associated SNP set, AS\_cross\_disease\_114, more cell types were detected, including B cells, naïve primary T cells and CD4 memory primary cells, with increased levels of significance (Supplementary Figure S1 and Supplementary Table S1).

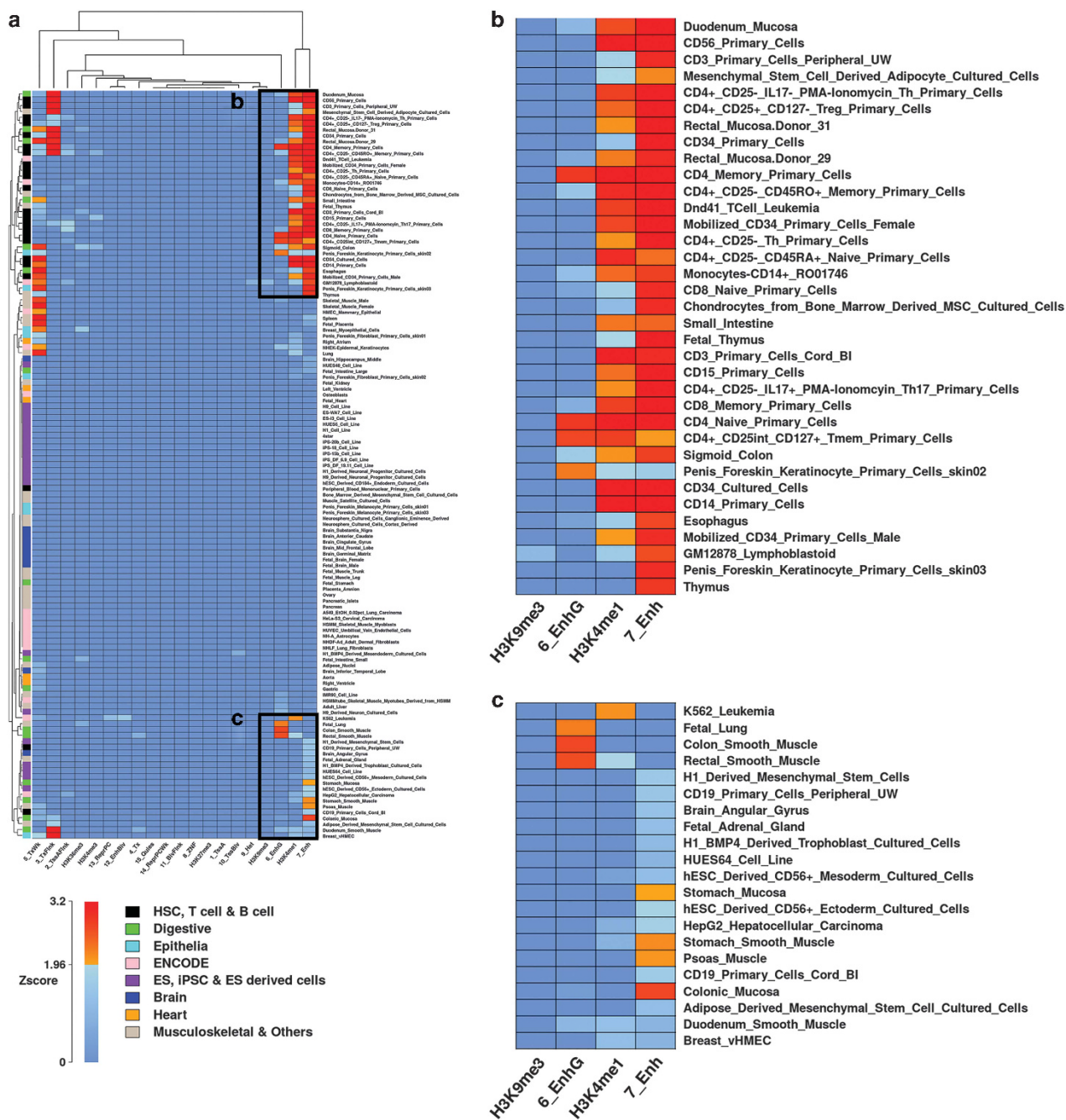
As histone marks such as H3K27ac and H3K4me1 can indicate nearby active enhancers,<sup>33</sup> we further extended the epigenetic analysis by utilizing the consolidated NIH Roadmap Epigenomics database. We focussed on the 20 epigenetic marks with data in all 127 cell types in the database. AS GWS-associated SNPs from the AS\_GWS\_40 set significantly colocalized with H3K4me1 and enhancers (7\_Enh) in T cells, monocytes, CD4 naïve T cells, T helper naïve cells and NK cells (Figure 2). Similar enrichment patterns were observed in both H3K4me1 and 7\_Enh, and these two epigenetic marks were clustered together (Figure 2 and Supplementary Table S2). In addition to immune cells, AS\_GWS\_40 SNPs also showed significant enrichment in enhancer-related epigenetic regions in gut cells (small intestine, rectal mucosa, duodenal mucosa and sigmoid colon). A relatively weak enrichment in B cells was observed in genic enhancers. The signal for gut cell types increased for the larger SNP set, AS\_cross\_disease\_114 (Supplementary Figure S2 and Supplementary Table S2). Furthermore, enrichment signal ( $P$ -adjust = 0.0025) was observed for bone marrow-derived chondrocytes in 7\_Enh enhancers with the AS\_GWS\_40 SNPs. Notably, the bone cell signal is much weaker than that seen in immune cells in AS\_GWS\_40 SNPs. Both chondrocytes and osteoblasts showed significant signals with the AS\_cross\_disease\_114 SNP set (Supplementary Figure S2).

AS-associated loci were enriched in gut and immune cell types and their response to cytokines

As many intergenic SNPs have *cis*-acting effects on gene expression, we examined cell type-specific expression of genes in proximity to AS-associated SNPs. To this end, we employed the SNPsea program<sup>34</sup> to identify potential cell types and pathways affected against four public gene expression databases. AS\_GWS\_40 loci were significantly enriched in gut and immune cell types, especially in monocytes (Table 1 and Supplementary Table S3). Similar cell types identified in previous epigenetics analyses, including CD4<sup>+</sup> memory T cells and gut cells, were also identified here. Gene Ontology (GO) enrichment analysis showed



**Figure 1.** Cell type specificity of AS\_GWS\_40 SNP set with H3K4me3 chromatin marks using NIH Roadmap cell types data using EPIGWAS. AS GWS-associated SNPs are enriched in association specifically with immune cell types and rectal mucosa. The blue line indicates  $P$ -value = 0.05, and the red line,  $P$ -value = 0.01.



**Figure 2.** AS-associated SNPs show enrichment in enhancer-related epigenetic regions in immune cells and gut cells. **(a)** Heat map showing epigenetic marks and chromatin state enrichment of AS\_GWS\_40 SNP set within the NIH Roadmap and ENCODE cell types. **(b, c)** Enlarged views of the boxed areas of **b** and **c** (**a**) respectively. Note that in addition to immune and gut cells, AS-associated SNPs also showed a relatively weak enrichment in B cells. Orange to red indicates significance, whereas cornflower blue to light blue indicates nonsignificance. Cell and tissue are clustered into eight groups: haematopoietic stem cells (HSCs), T cells and B cells; digestive cells; epithelial cells; cell lines from ENCODE; embryonic stem (ES) cells, induced pluripotent stem cells (iPSCs) and ES-derived cells; brain cells; heart cells; musculoskeletal and other cells, and coloured according to the key. For a full listing of the cell types and chromatin states in **(a)**, see Supplementary Table S15.

that these AS\_GWS\_40 loci were predicted to be involved in pathways responding to cytokine stimulus (Table 1), consistent with reports that the cytokines interleukin (IL)-17 and IL-23 play important roles in the pathogenesis of AS.<sup>35,36</sup> The AS\_cross\_disease\_114 SNP set was analysed in a similar way and were enriched in immune cell types, including monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> cells and their subsets (Supplementary Table S3).

AS differentially expressed genes showed enrichment in different immune cell types

Genes demonstrated by our in-house RNA sequencing (RNA-Seq) to be differentially expressed in peripheral blood mononuclear cells between AS cases and healthy controls (herein termed 'AS differentially expressed genes') were then analysed for the cell type in which they were primarily expressed (a listing of the AS

**Table 1.** AS-associated loci were enriched in gut and immune cell types, particularly in monocytes, and involved in cytokine response pathways

Database	Identified cell types and biological processes
NIH	CD4 <sup>+</sup> memory, sigmoid colon
BLUEPRINT	Monocytes (adjusted <i>P</i> -value = 0.072)
FANTOM5	Monocytes, migratory Langerhans cells, eosinophils, basophils
GO	Response to cytokine stimulus, cellular response to cytokine stimulus

Abbreviations: AS, ankylosing spondylitis; GO, Gene Ontology; NIH, National Institutes of Health. Cell types and biological processes were identified using AS-associated loci from the AS\_GWS\_40 single-nucleotide polymorphism (SNP) set with SNPsea analyses and GO enrichment analysis.

differentially expressed genes appear in Supplementary Table S4). This was performed by merging the NIH Roadmap and BLUEPRINT RNA-Seq data to form a combined data set representing RNA-Seq data from 93 cell types. As with epigenetic studies, and studies of genes near AS-associated SNPs, the primary cell types that demonstrated enrichment of expression of AS differentially expressed genes were circulating immune cells (Figure 3 and Supplementary Table S5). This was expected, particularly given the fact that the in-house RNA-Seq data are from RNA from peripheral blood mononuclear cells. No enrichment of expression was noted in intestinal or other noncirculating tissues other than in breast myoepithelial. AS upregulated genes (from 1.2- to 1.4-fold) showed different enrichment patterns compared with AS downregulated genes (0.9-fold; Figure 3). AS upregulated genes were most enriched in monocytes (adjusted *P*-value = 0.00093), whereas downregulated genes were most enriched in CD8-naïve primary cells compared with all the human genes in the RNA-Seq data set. Only the mixed peripheral blood mononuclear cell types showed both up- and down-regulated genes. AS upregulated genes also showed significant enrichment in mature neutrophil, macrophage and proinflammatory macrophages after multiple hypothesis testing correction. AS downregulated genes showed relatively weak, but significant, enrichment in CD4<sup>+</sup> primary T cells, CD56<sup>dim</sup> NK cells, central/effector CD8<sup>+</sup> T cells, central/effector CD4<sup>+</sup> T cells and regulatory T cells. Bone cells were not represented in the NIH Roadmap and BLUEPRINT RNA-Seq databases, and thus were not able to be tested in this analysis. A high degree of correlation of gene expression patterns was observed between cell types, reducing the power of this analysis to resolve cell type-specific expression patterns of AS differentially expressed genes.

To determine whether the differentially expressed gene set also showed cell type bias in protein expression levels, the E-PROT-1 data set from Expression Atlas was assessed using a similar methodology to that employed for RNA-Seq data analyses. A very similar pattern of protein expression was observed as for RNA expression (Figure 4 and Supplementary Table S5). Proteins encoded by AS upregulated genes showed significant enrichment in different cell types versus AS downregulated ones. Proteins encoded by AS upregulated genes showed the most enrichment in monocytes, whereas those encoded by AS downregulated genes showed enrichment in CD8<sup>+</sup> T cells and NK cells. Proteins encoded by AS upregulated genes also showed significant enrichment in gastrointestinal tissues, including rectum and colon.

AS gut consensus loci enriched in pathway implicating microbes and the gut

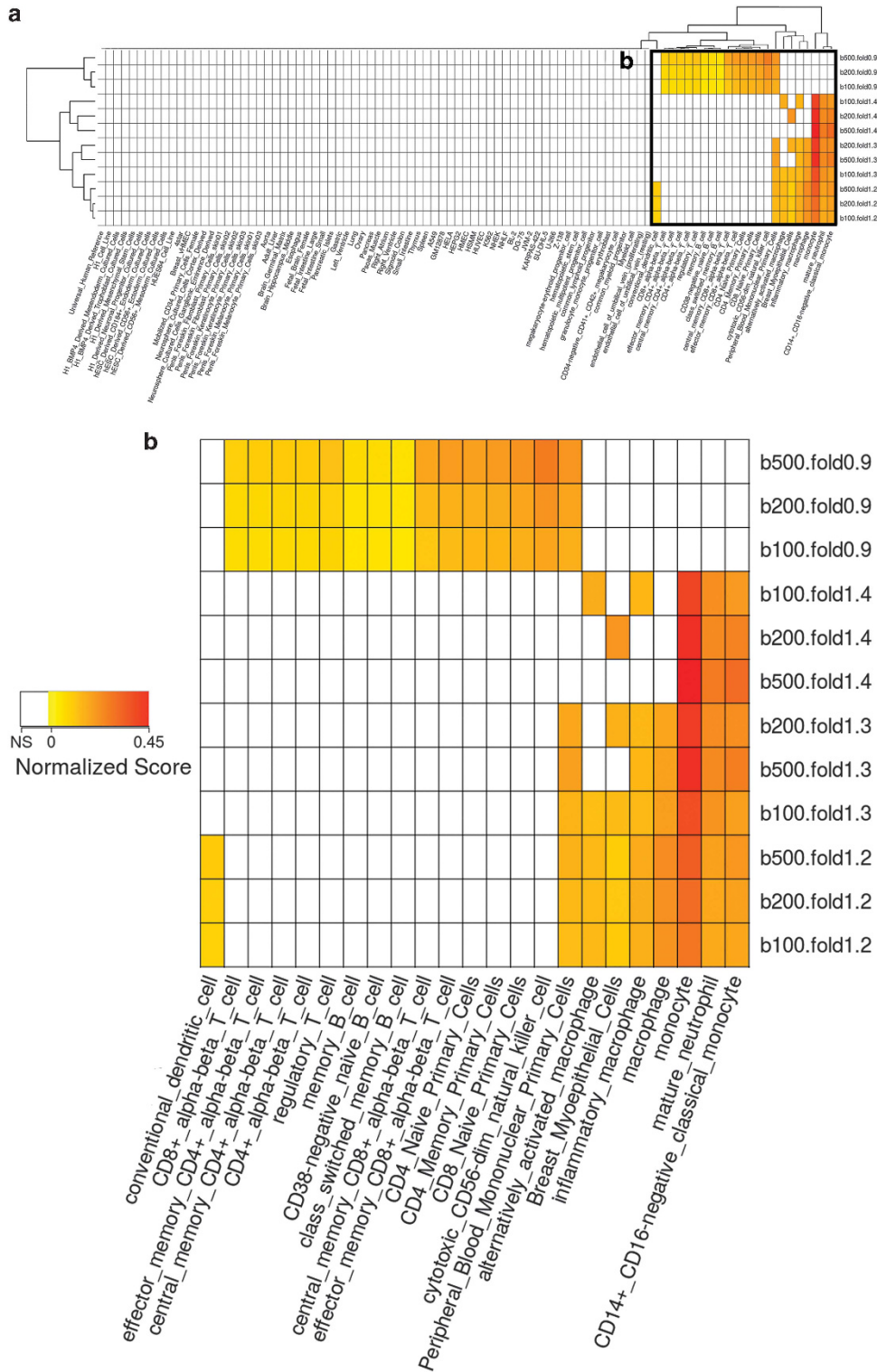
Overall, the AS\_GWS\_40 SNPs were significantly enriched in nine gut cell types in 7\_Enh enhancer regions. To dissect this further, we defined SNPs in 7\_Enh enhancer regions with a locus score lower than 1 (see Materials and methods) in at least six gut cell

types cells as consensus gut SNPs, and analysed these in more depth. DAVID (the Database for Annotation, Visualization and Integrated Discovery) annotation (version 6.8) with high classification stringency was employed to perform GO enrichment analysis. There were 14 gut consensus SNPs that were located in 13 loci (all 13 loci were matched with DAVID IDs). Enrichment analysis identified multiple pathways that implicate microbes and the gut; GO term 'response to bacterium', with 7 out of 13 loci being involved (*TNFRSF1A*, *LTBR*, *IL27*, *ERAP1*, *NFKB1*, *IL6R* and *NOS2*, *P*-value =  $3.3 \times 10^{-7}$ ; Supplementary Table S6), and GO term 'response to external biotic stimulus', 'response to other organism' and 'response to biotic stimulus' with the same 7 loci involved (*P*-value =  $6.6 \times 10^{-6}$ ,  $6.6 \times 10^{-6}$  and  $8.5 \times 10^{-6}$ , respectively). Out of the 13 AS loci, 5 were also GWS IBD associated (*C1orf106*, *IL1RL2*, *IL27*, *NFKB* and *LTBR*).

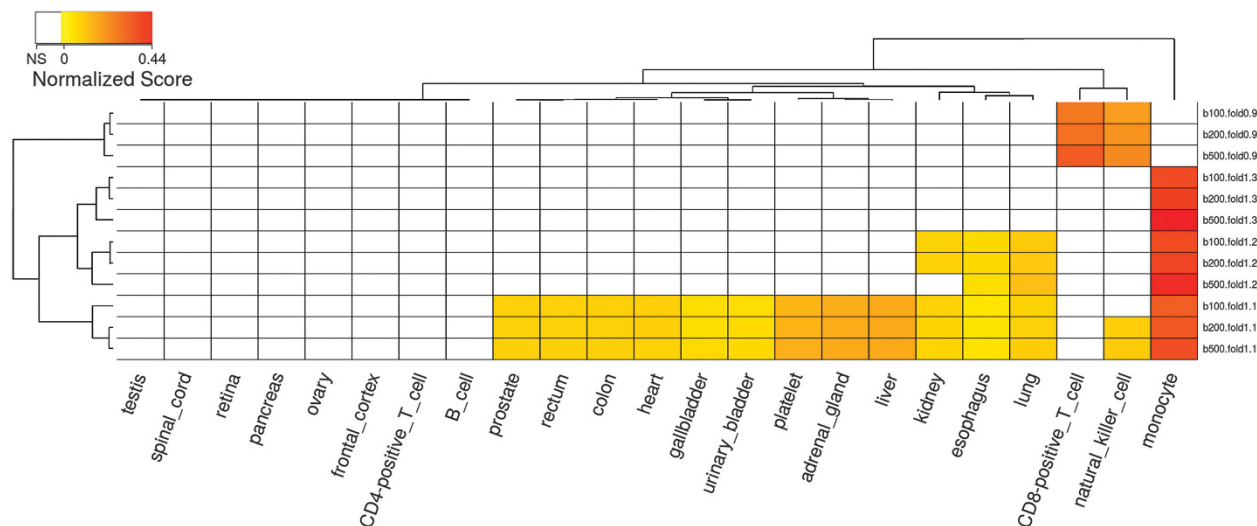
Two more gastrointestinal cells (fetal intestine large and fetal intestine small) were identified using the AS\_cross\_disease\_114 SNP set in enhancer regions (7\_Enh, 10 gut cells). A similar GO term enrichment analysis was performed on this SNP set that is enriched for IBD genes (70 GWS IBD-associated SNPs out of 113 non-MHC SNPs). SNPs in 7\_Enh enhance regions with a locus score lower than 1 in at least six gut cells were considered as consensus gut SNPs. There were 43 gut consensus SNPs that were located in 41 loci (36/41 are matched in DAVID), and enrichment analysis identified multiple pathways (Supplementary Table S7); GO term 'response to bacterium', with 13 of 36 loci being involved (*TNFRSF1A*, *FOS*, *CARD9*, *LTBR*, *PTGER4*, *IRF5*, *IL27*, *ERAP1*, *NFKB1*, *IL6R*, *NOS2*, *IL10* and *TNFSF8*, *P*-value =  $3.6 \times 10^{-10}$ ), GO term 'response to biotic stimulus', with 16 loci involved (*CARD9*, *LTBR*, *PTGER4*, *IL27*, *NFKB1*, *IL6R*, *IL10*, *TNFSF8*, *LSM14A*, *FOS*, *TNFRSF1A*, *IRF5*, *ERAP1*, *NOS2*, *PPP2R3C* and *CD28*, *P*-value =  $1.1 \times 10^{-5}$ ) and GO term 'response to other organism' and 'response to external biotic stimulus', with 15 loci involved (*CARD9*, *LTBR*, *PTGER4*, *IL27*, *NFKB1*, *IL6R*, *IL10*, *TNFSF8*, *LSM14A*, *FOS*, *TNFRSF1A*, *IRF5*, *ERAP1*, *NOS2* and *CD28*, *P*-values are both  $8.4 \times 10^{-10}$ ). There were eight IBD GWS loci (*CARD9*, *FOS*, *IL10*, *IL27*, *IRF5*, *LTBR*, *NFKB1* and *TNFSF8*) and seven AS GWS loci (*ERAP1*, *IL27*, *IL6R*, *LTBR*, *NFKB1*, *NOS2* and *TNFRSF1A*) among these genes. This census SNP set shared 26 loci (26 SNPs, 26 loci, matched 21 DAVID IDs) that are IBD GWS, whereas 16 out of 41 loci (18 SNPs, 16 loci, matched 16 DAVID IDs) are AS GWS loci. The 26 IBD GWS shared loci identified an enrichment in GO terms 'response to bacterium', 'response to other organism', 'response to external biotic stimulus' and 'response to biotic stimulus' (Supplementary Table S8), whereas the 16 AS GWS loci were enriched in GO terms 'response to bacterium', 'response to external biotic stimulus', 'response to other organism', 'response to biotic stimulus' and 'immune response' (Supplementary Table S9).

## DISCUSSION

Now that large numbers of robust genetic associations have been identified for common diseases, a major challenge is faced to identify the causative variants and unravel the mechanisms through which they operate. This is particularly challenging in the case of intergenic variants that make up ~65–72% of the known AS-associated variants, a similar proportion to other immune-mediated diseases. Furthermore, causal variants in LD with disease-associated variants may affect a limited number of cell types or tissues. Therefore, in the present study we leveraged cell type-specific epigenetic, gene expression and protein expression data to identify the cell types involved in AS pathogenesis using a larger AS-associated SNP set recently identified in a large cohort cross-disease study.<sup>8</sup> A limitation of this kind of approach is the lack of epigenetic annotation and gene expression data for some AS-relevant tissues, including enthesial and synovial cell populations, and a paucity of data from relevant intestinal cell subsets, such as innate lymphoid cells and Paneth cells, that have



**Figure 3.** AS upregulated and downregulated genes were enriched in different blood cell types. **(a)** Heat map showing expression of AS differentially expressed genes based on RNA-Seq data. **(b)** Enlarged view of the boxed area *b* of **(a)**. Note that different cell types were identified based on either AS up-regulated or down-regulated genes. The enrichment scores were normalized by the number of genes. Yellow to red colours indicate significance after Bonferroni correction ( $\alpha=0.05$ ), whereas white colour indicates no significance (NS). Normalization, reads per kilobase per million. For a full listing of the cell types in **(a)**, see Supplementary Table S15. Differentially expressed genes were grouped by different basemean counts (mean of normalized counts of all samples) as an output from DESeq2, and labelled as b100, b200 and b500, where b100 represents genes with basemean counts of  $\geq 100$  and so on. The fold change is also indicated.



**Figure 4.** AS up-regulated and down-regulated genes were enriched in different blood cell types, as well as gut and other cell types, in proteomics data. Heat map showing expression of AS differentially expressed genes based on proteomics data. The enrichment scores were normalized by the number of genes. Yellow to red colours indicate significance after Bonferroni correction ( $\alpha = 0.05$ ), whereas white colour indicates no significance (NS). Basemean and fold change are as described in Figure 3.

previously been implicated in AS pathogenesis. Nevertheless, the data presented here indicate that monocytes, CD4<sup>+</sup>, CD8<sup>+</sup> T cells, NK cells, regulatory T cells and gut cells are relevant cell types in AS, in reasonable agreement with a recent study that used a similar epigenetic analysis with predicted causal SNPs (using PICS) that identified CD4<sup>+</sup> T cells, Th1, Th17, monocytes and intestinal mucosal cells as being implicated in AS.<sup>30</sup>

Despite the strength of association with *HLA-B27* and, more recently, components of class I antigen processing machinery such as *ERAP1*, *ERAP2* and *NPEPPS*, the precise role of the HLA class I system of antigen presentation in AS pathogenesis remains unclear. Several genes involved in the regulation of CD8<sup>+</sup> T-cell number and function are GWS associated with AS, including *EOMES*, *IL7R*, *RUNX3* and *TBX21* (that encodes the transcription factor, T-bet).<sup>11</sup> Total CD8<sup>+</sup> T-cell counts are lower in the periphery of AS patients compared with healthy controls,<sup>12</sup> although numbers of cytotoxic, perforin-secreting CD8<sup>+</sup> T cells are elevated in AS cases.<sup>37</sup> Recently, carriage of risk variants in *TBX21* has been shown to expand a population of IFN- $\gamma$ <sup>+</sup> IL-17<sup>+</sup> inflammatory cells in the circulation of AS patients.<sup>38</sup> However, the coordinated effects of IL-7R, *EOMES*, *RUNX3* and T-bet on CD8<sup>+</sup> T-cell function in AS remain to be determined.

Enrichment of AS-associated SNPs identified enhancer and gene expression data related to CD4<sup>+</sup> memory T cells, with nine of these loci (*RP11-84D1.2*, *IL6R*, *C1orf106*, *GPR65*, *NOS2*, *ERN1*, *FAM118A*, *NFKB1* and *ERAP1*) shared in both epigenetic and gene expression enrichment analyses. Nuclear factor- $\kappa$ B plays an important role in regulating different CD4<sup>+</sup> T-cell subsets<sup>39</sup> among other diverse immunological functions, and IL-6 is an important mediator of the immune response especially by directly acting on CD4<sup>+</sup> T cells.<sup>40</sup>

Although the primary MHC associations of AS is with HLA class I alleles such as HLA-B27, the disease is also associated with HLA class II alleles.<sup>13</sup> CD4<sup>+</sup> T cells, which classically recognize HLA class II-presented antigens, have been shown to be an important source of IL-17 in AS cases.<sup>22</sup> CD4<sup>+</sup> T cells expressing the killer cell immunoglobulin receptor, KIR3DL2, recognize HLA-B27 homodimers, and this recognition is associated with the secretion of large amounts of inflammatory cytokines.<sup>41</sup> Genes such as *ICOSLG*, *BACH2* and *SH2B3* that regulate CD4<sup>+</sup> T-cell function have also been associated with AS.<sup>11</sup> The classic definition of a Th17 cells is that of a CD4<sup>+</sup> T-cell secreting IL-17. In this regard, CD4<sup>+</sup> T-cell function may be heavily influenced by polymorphisms in

components of the IL-23 signalling pathway, including *IL23R*, *IL27*, *IL12B*, *JAK2*, and *TYK2*, that have all been linked with increased AS susceptibility.<sup>8–11</sup> Our analyses here provide further evidence of the involvement of this cell type in AS pathogenesis.

B cells were identified on the enrichment of AS\_cross\_disease\_114 SNPs in enhancer-related epigenetic regions. B cells were further implicated by AS downregulated genes being enriched in these cells (in contrast to the AS genes upregulated in T cells; Figure 3). Although there is some evidence that the B cell-depleting agent, rituximab, may be beneficial for AS patients, there appears to be variability in response to that agent with many patients being nonresponders.<sup>42</sup> It remains to be determined whether individuals who responded to rituximab carry the SNPs that were enriched in B-cell signals.

The role of innate cell types has been largely overshadowed by the extensive efforts to characterize T-cell responses in AS. Our data suggest that NK cells and monocytes are important regulators of inflammatory processes in AS, and AS patient-derived monocytes express a range of inflammatory proteins.<sup>43</sup> It is yet to be determined how monocytes contribute to AS pathogenesis, although they may present microbial antigens derived from an altered gut microbiome, or they may be an important source of inflammatory cytokines, such as IL-23. Indeed, monocytes have been shown to secrete inflammatory cytokines in axial SpA.<sup>44</sup> Similarly, potential roles of NK cells in AS are not well described. In a simple model, NK cells may operate in AS through HLA-KIR interactions.<sup>23</sup> We have recently shown that NK cells expressing the transcription factor T-bet are potent producers of IL-17 and interferon- $\gamma$ .<sup>38</sup> However, NK cells may also play important protective roles at certain tissue sites, including the gut where they secrete immunoprotective IL-22.<sup>45</sup> The data presented here should focus research on understanding how CD4<sup>+</sup> and CD8<sup>+</sup> T cells, NK cells and monocytes operate in AS pathogenesis. Profiling the transcriptome of these cell types in large numbers of cases and controls would aid in understand how these cell types are altered in disease and how these cell types operate at sites of inflammation in AS.

Microbial interaction with the gut has been suggested in AS, and there is a discrete microbial signature in the terminal ileum of patients with AS compared with healthy subjects.<sup>29</sup> The microbial composition correlated with disease status and greater differences were observed between disease groups than within disease groups. Furthermore, AS-associated genes play a role in gut

immunity, particularly genes involved in the IL-23 pathway (reviewed in refs 46,47). In the present study, the AS\_GWS\_40 SNP set showed enrichment in gastrointestinal cells in addition to immune cells. Our enrichment analyses identified GO term pathways, some implicating microbes and the gut; 'defense response', 'response to bacterium', 'response to molecule of bacterial origin', 'defense response to bacterium', 'immune response' and 'innate immune response'. This is also in reasonable agreement with a study that applied pathway enrichment analysis with 30 AS-associated genes that identified GO term pathways 'cytokine production' and 'response to bacterium' as the most relevant to AS.<sup>31</sup> One likely factor contributing to the difference between the studies was that we analysed up- and down-regulated AS genes. Of the pathways we identified, a number of involved loci are associated with AS; *NOS2* and *CARD9* are known to be involved in the regulation of IgA function and the composition of the gut microbiota, and to maintain the balance between the host and the microbiota, whereas IL-10 plays a role in the development of intestinal immunity.<sup>48</sup> Together with the results of the AS\_GWS\_40 SNPs set, they indicate that not only IBD genes but also AS genes were enriched in gastrointestinal tissue. These results further implicate microbes and the gut in the aetiology of AS and highlight the overlap in genetic risk for IBD and AS. This provides further support to the hypothesis that AS is caused by interactions between the host gut immune system and the gut microbiome.<sup>49,50</sup>

The relatively weak enrichment signals for bone cell types are consistent with bone disease being a secondary manifestation, not a pathogenic driver for AS. There are thought to be two disease phases in AS: an initial immune-driven inflammatory period that then progresses, to varying degrees, to the debilitating joint fusion that characterizes the disease, with the changes in the joints being secondary to the initial inflammatory changes.<sup>51–53</sup> Unfortunately, there were no bone cell data available in gene expression or protein expression datasets for our analyses. Therefore, it is difficult to investigate whether AS-associated genes were differentially expressed in bone cell types. Nevertheless, changes in bone-specific genes would be likely to alter disease progression and severity rather than initiate AS pathogenesis.

Discriminating pathogenic from bystander inflammatory cell subsets has proven difficult in AS, with many genomic and functional studies implicating a wide range of cell types.<sup>8,54,55</sup> This study shows that AS-associated variants primarily operate through effects in CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, NK cells and monocytes, as well as gastrointestinal cells. It also highlights that AS pathogenesis is affected by the relationship between the host genetics, the intestinal microbiome and the immune response. As the number of cell type-specific databases increase, including epigenetics and gene expression data, further integration of multiomics data will continue to advance our understanding of AS pathogenesis.

## MATERIALS AND METHODS

### SNP sets

So far, 39 non-MHC susceptibility variants showed GWS in AS-only analyses. However, a further 74 non-MHC GWS variants were identified in analyses leveraging pleiotropy with other related diseases (Crohn's disease, psoriasis, primary sclerosing cholangitis and ulcerative colitis).<sup>8</sup> This SNP set of 113 SNPs with the HLA-B27 tag SNP, rs116488202, was labelled 'AS\_cross\_disease\_114', and the set of 40 AS-only GWS SNPs set, 'AS\_GWS\_40'. The complete list of SNP IDs used appears in Supplementary Table S10, and a summary of analyses performed in this study appears in Supplementary Table S11. All coordinates used in this study were converted using the UCSC (University of California, Santa Cruz) Genome Browser liftOver tool to the GRCh37/hg19 assembly.

### Epigenetic marks data set

A total of 127 consolidated epigenomes were generated from the integrated study of 111 reference human epigenomes in NIH Roadmap Epigenomics Mapping Consortium (Release 9 of the compendium) and 16 cell lines in ENCODE. Each of the 127 epigenomes contains chromatin immunoprecipitation sequencing data sets for a core set of histone modifications (H3K4me1, H3K4me3, H3K27me3, H3K36me3 and H3K9me3; for the putative functions of the histone modifications, see Supplementary Table S12). Epigenetic mark data utilized included all processed narrowPeak data of histone modifications, chromatin immunoprecipitation sequencing, DNase-seq (processed by MACS 2), reduced representation bisulfite sequencing data and the putative core 15 chromatin states data set for all 127 epigenomes (Supplementary Table S13). Data were available for all of these marks with data in 127 cell types (Supplementary Table S14).

### Gene and protein expression data sets

Three gene expression data sets were used for differentially expressed gene set enrichment analysis, including FANTOM5 CAGE human expression data (533 cell types), RNA-Seq data from integrated NIH Roadmap compendium (57 cell types) and RNA-Seq data from BLUEPRINT of Haematopoietic Epigenomes project (36 cell types). We summed the CAGE entries for each gene and averaged them. The protein data set used was the Expression Atlas (<https://www.ebi.ac.uk/gxa/experiments/E-PROT-1>),<sup>56</sup> a data set that contains protein expression level information for 23 cell/tissue types.

### Identification of the AS differentially expressed gene set

The set of genes differentially expressed between AS patients and healthy controls was derived from an in-house RNA-Seq study that will be the subject of a separate report. This study was undertaken in accordance with the Australian National Statement on Ethical Conduct in Human Research (National Health and Medical Research Council of Australia) approved by the Metro South Health Service District Human Research Ethics Committee (HREC/05/QPAH/221) and The University of Queensland Medical Research Ethics Committee (reference 2006000102). All individuals involved gave written informed consent. The 69 patients who had met the modified New York criteria for AS and 79 healthy controls, all of whom were Caucasian, were involved in the study. AS patients who were taking anti-TNF treatment were excluded from this study. The mean age of AS patients was 39.8 ± 11.6 years, and of healthy controls, 40.9 ± 11.2 years. Among the AS patients, the mean erythrocyte sedimentation rate was 22.2 ± 21.5 mm h<sup>-1</sup> (median, 13 mm h<sup>-1</sup>), mean C-reactive protein level was 16.0 ± 16.2 mg l<sup>-1</sup> (median, 9.2 mg l<sup>-1</sup>) and mean Bath Ankylosing Spondylitis Disease Activity Index was 5.3 ± 2.6 (median, 5.8).

Peripheral blood mononuclear cell samples were collected and 1 µg RNA from each sample was used for library preparation using the Illumina TruSeq Stranded Total RNA with Ribo-Zero for Human/Mouse/Rat Low-throughput sample prep kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Samples were pooled into groups of 12. Flow cells were sequenced using reagents from TruSeq SBS kit v3 (Illumina) on either a HiSeq2000 or a HiSeq 2500 (Illumina) running in high output mode. Read quality was assessed using FastQC in each sequencing lane. Sequenced reads were mapped back to the human genome reference hg19 using TopHat version 2.0.6 (<http://ccb.jhu.edu/software/tophat/index.shtml>). The table of counts created by mapping reads was generated by HTSeq. Data were assessed for systemic differences caused by factors other than patient status using principal components analysis from the lattice package for R (<https://cran.r-project.org/>). Only genes that had at least 10 reads mapped to them in at least 15 samples were included in the analysis (20 088 genes). Differential gene expression was determined using DESeq2.<sup>57</sup>

### Cell type specificity score with an epigenetic mark

To compare the extent of enrichment of SNP location within epigenetic marks in specific cell types, we calculated a cell type specificity score. For each testing SNP set, *i*, and epigenetic mark, *m*, in cell type, *c*, we calculated the cell type specificity score,

$$S_{i,m,c} = \frac{n_{i,m,c}}{N_i} \quad (1)$$

where  $N_i$  is the number of SNPs in the SNP set  $i$ , and  $n_{i,m,c}$  is the number of SNPs overlapping with the epigenetic mark,  $m$ , in cell type,  $c$ . A SNP was considered to be overlapping with the epigenetic mark,  $m$ , if any other SNP in LD ( $r^2 > 0.8$ ) with the original SNP was located within the mark narrowPeak region.

#### Assessment of SNP enrichment in biological annotations

In order to calculate enrichment in biological annotations within the AS cross-disease SNP sets, we accounted for potential biases by preparing a pool of appropriately matched background SNPs for comparison with the query SNPs based on GC content, minor allele frequency, distance to nearest transcription start site and number of SNPs in LD with the testing SNP. GC content was calculated within a 101 bp window extending 50 bps both sides of the testing SNP. European ancestry minor allele frequencies were extracted from the 1000 Genome Project Phase1 Integrated v3 call data set. The transcription start site locations were defined by GENCODE Project (v19). The number of SNPs in LD ( $r^2 > 0.8$ ) with the query SNP was calculated from the European ancestry data from the 1000 Genome Project Phase1 Integrated v3 call data set. SNPs were grouped into different bins according to GC content, minor allele frequency, transcription start site and number of LD SNPs. If no matched background SNP could be identified for the query SNP, SNPs in neighbouring bins were considered. A permutation test and multiple testing correction was then performed thus: a null distribution was constructed by 100 000 permutations. Each time, a random SNP set with the same number of SNPs was extracted from the matched background SNP sets, and the cell type specificity score was calculated under the same epigenetic mark. As described in equation (2), the empirical  $P$ -value was calculated as the adjusted proportion of permutations for which the cell type specificity score is equal or greater than the query SNP set;<sup>58</sup>

$$p = 1 + x_{m,c}/1 + X_{m,c} \quad (2)$$

where  $X_{m,c}$  is the total number of permutations (100 000) and  $x_{m,c}$  is the number of permutations for which the cell type specificity score is equal or greater than the query SNP set. Bonferroni correction was employed to correct for multiple testing.

#### Locus score

In order to rank the informative loci, a locus score was calculated. Similar to the cell type specificity score, the locus score was calculated as the number of matched random SNPs overlapping with epigenetics region divided by the total number of matched random SNPs. The locus score is the probability of a SNP (with the restraints of minor allele frequency, GC content, transcription start site and number of LD SNPs) overlapping an epigenetics marker region by chance. The lower the score, the more significantly the target SNP overlaps with an epigenetic mark. Locus scores  $< 1$  were considered meaningful. A gene was assigned to the SNP by the nearest distance to the gene region defined in Ensembl database (GRCh37, Release 75).

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank the participating patients and healthy controls for taking part in this study. MAB is funded by a National Health and Medical Research Council Senior Principal Research Fellowship.

#### REFERENCES

- Mielants H, Veys EM, Cuvelier C, De Vos M, Botelberghe L. HLA-B27 related arthritis and bowel inflammation. Part 2. Ileocolonoscopy and bowel histology in patients with HLA-B27 related arthritis. *J Rheumatol* 1985; **12**: 294–298.
- Brown MA, Kennedy LG, MacGregor AJ, Darke C, Duncan E, Shatford JL et al. Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. *Arthritis Rheum* 1997; **40**: 1823–1828.
- Brown MA, Laval SH, Brophy S, Calin A. Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. *Ann Rheum Dis* 2000; **59**: 883–886.

- Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A et al. Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. *Arthritis Rheum* 1998; **41**: 58–67.
- Zeng QY, Chen R, Darmawan J, Xiao ZY, Chen SB, Wigley R et al. Rheumatic diseases in China. *Arthritis Res Ther* 2008; **10**: R17.
- Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, Sturrock RD. Ankylosing spondylitis and HL-A 27. *Lancet* 1973; **1**: 904–907.
- Caffrey MF, James DC. Human lymphocyte antigen association with ankylosing spondylitis. *Nature* 1973; **9**: 121.
- Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016; **48**: 510–518.
- Wellcome Trust Case Control Consortium/Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, Craddock N et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007; **39**: 1329–1337.
- Danoy P, Pryce K, Hadler J, Bradbury LA, Farrar C, Pointon J et al. Association of variants at 1q32 and STAT3 with ankylosing spondylitis suggests genetic overlap with Crohn's disease. *PLoS Genet* 2010; **6**: e1001195.
- International Genetics of Ankylosing Spondylitis Consortium (IGAS), Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 2013; **45**: 730–738.
- Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 2011; **43**: 761–767.
- Cortes A, Pulit SL, Leo PJ, Pointon JJ, Robinson PC, Weisman MH et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. *Nat Commun* 2015; **6**: 7146.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190–1195.
- Kumar V, Westra HJ, Karjalainen J, Zernakova DV, Esko T, Hrdlickova B et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 2013; **9**: e1003201.
- Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010; **42**: 1118–1125.
- Fraser HB, Xie XH. Common polymorphic transcript variation in human disease. *Genome Res* 2009; **19**: 567–575.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 2010; **28**: 1045–1048.
- Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015; **47**: 381–386.
- Szalay B, Meszaros G, Cseh A, Acs L, Deak M, Kovacs L et al. Adaptive immunity in ankylosing spondylitis: phenotype and functional alterations of T-cells before and during infliximab therapy. *Clin Dev Immunol* 2012; **2012**: 808724.
- Shen H, Goodall JC, Hill Gaston JS. Frequency and phenotype of peripheral blood Th17 cells in ankylosing spondylitis and rheumatoid arthritis. *Arthritis Rheum* 2009; **60**: 1647–1656.
- Chan AT, Kollnberger SD, Wedderburn LR, Bowness P. Expansion and enhanced survival of natural killer cells expressing the killer immunoglobulin-like receptor KIR3DL2 in spondylarthritis. *Arthritis Rheum* 2005; **52**: 3586–3595.
- Kenna TJ, Davidson SJ, Duan R, Bradbury LA, McFarlane J, Smith M et al. Enrichment of circulating interleukin-17-secreting interleukin-23 receptor-positive gamma/delta T cells in patients with active ankylosing spondylitis. *Arthritis Rheum* 2012; **64**: 1420–1429.
- Niu XY, Zhang HY, Liu YJ, Zhao D, Shan YX, Jiang YF. Peripheral B-cell activation and exhaustion markers in patients with ankylosing spondylitis. *Life Sci* 2013; **93**: 687–692.
- Noordenbos T, Yeremenko N, Goñta I, van de Sande M, Tak PP, Canete JD et al. Interleukin-17-positive mast cells contribute to synovial inflammation in spondylarthritis. *Arthritis Rheum* 2012; **64**: 99–109.
- Ciccio F, Bombardieri M, Principato A, Giardina A, Tripodo C, Porcasi R et al. Overexpression of interleukin-23, but not interleukin-17, as an immunologic signature of subclinical intestinal inflammation in ankylosing spondylitis. *Arthritis Rheum* 2009; **60**: 955–965.
- Ciccio F, Bombardieri M, Rizzo A, Principato A, Giardina AR, Raiata F et al. Over-expression of Paneth cell-derived anti-microbial peptides in the gut of



- patients with ankylosing spondylitis and subclinical intestinal inflammation. *Rheumatology (Oxford)* 2010; **49**: 2076–2083.
- 29 Costello ME, Ciccio F, Willner D, Warrington N, Robinson PC, Gardiner B *et al*. Intestinal dysbiosis in ankylosing spondylitis. *Arthritis Rheumatol* 2014; **67**: 686–691.
- 30 Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S *et al*. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015; **518**: 337–343.
- 31 Uddin M, Codner D, Hasan SM, Scherer SW, O'Rielly DD, Rahman P. Integrated genomics identifies convergence of ankylosing spondylitis with global immune mediated disease pathways. *Sci Rep* 2015; **5**: 10314.
- 32 Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS *et al*. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013; **45**: 124–130.
- 33 Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 2014; **15**: 272–286.
- 34 Slowikowski K, Hu XL, Raychaudhuri S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* 2014; **30**: 2496–2497.
- 35 Wang X, Lin Z, Wei Q, Jiang Y, Gu J. Expression of IL-23 and IL-17 and effect of IL-23 on IL-17 production in ankylosing spondylitis. *Rheumatol Int* 2009; **29**: 1343–1347.
- 36 Mei Y, Pan F, Gao J, Ge R, Duan Z, Zeng Z *et al*. Increased serum IL-17 and IL-23 in the patient with ankylosing spondylitis. *Clin Rheumatol* 2011; **30**: 269–273.
- 37 Schirmer M, Goldberger C, Wurzner R, Duftner C, Pfeiffer KP, Clausen J *et al*. Circulating cytotoxic CD8(+) CD28(-) T cells in ankylosing spondylitis. *Arthritis Res* 2002; **4**: 71–76.
- 38 Lau MC, Keith P, Costello ME, Bradbury LA, Hollis KA, Thomas R *et al*. Genetic association of ankylosing spondylitis with TBX21 influences T-bet and pro-inflammatory cytokine expression in humans and SKG mice as a model of spondyloarthritis. *Ann Rheum Dis* 2016; **76**: 261–269.
- 39 Zhu J, Yamane H, Paul WE. Differentiation of effector CD4 T cell populations. *Annu Rev Immunol* 2010; **28**: 445–489.
- 40 Dienz O, Rincon M. The effects of IL-6 on CD4 T cell responses. *Clin Immunol* 2009; **130**: 27–33.
- 41 Bowness P, Ridley A, Shaw J, Chan AT, Wong-Baeza I, Fleming M *et al*. Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis. *J Immunol* 2011; **186**: 2672–2680.
- 42 Song IH, Heldmann F, Rudwaleit M, Listing J, Appel H, Braun J *et al*. Different response to rituximab in tumor necrosis factor blocker-naïve patients with active ankylosing spondylitis and in patients in whom tumor necrosis factor blockers have failed: a twenty-four-week clinical trial. *Arthritis Rheum* 2010; **62**: 1290–1297.
- 43 Wright C, Edelmann M, diGleria K, Kollnberger S, Kramer H, McGowan S *et al*. Ankylosing spondylitis monocytes show upregulation of proteins involved in inflammation and the ubiquitin proteasome pathway. *Ann Rheum Dis* 2009; **68**: 1626–1632.
- 44 Conrad K, Wu P, Sieper J, Syrbe U. In vivo pre-activation of monocytes in patients with axial spondyloarthritis. *Arthritis Res Ther* 2015; **17**: 179.
- 45 Ciccio F, Accardo-Palumbo A, Alessandro R, Rizzo A, Principe S, Peralta S *et al*. Interleukin-22 and interleukin-22-producing NKp44+ natural killer cells in sub-clinical gut inflammation in ankylosing spondylitis. *Arthritis Rheum* 2012; **64**: 1869–1878.
- 46 Costello ME, Elewaut D, Kenna TJ, Brown MA. Microbes, the gut and ankylosing spondylitis. *Arthritis Res Ther* 2013; **15**: 214.
- 47 Brown MA, Kenna T, Wordsworth BP. Genetics of ankylosing spondylitis—insights into pathogenesis. *Nat Rev Rheumatol* 2016; **12**: 81–91.
- 48 Kamada N, Seo SU, Chen GY, Nunez G. Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol* 2013; **13**: 321–335.
- 49 Kenna TJ, Brown MA. Immunopathogenesis of ankylosing spondylitis. *Int J Clin Rheumatol* 2013; **8**: 265–274.
- 50 Cua DJ, Sherlock JP. Autoimmunity's collateral damage: gut microbiota strikes 'back'. *Nat Med* 2011; **17**: 1055–1056.
- 51 Lories RJ, Schett G. Pathophysiology of new bone formation and ankylosis in spondyloarthritis. *Rheum Dis Clin North Am* 2012; **38**: 555–567.
- 52 Tseng HW, Pitt ME, Glant TT, McRae AF, Kenna TJ, Brown MA *et al*. Inflammation-driven bone formation in a mouse model of ankylosing spondylitis: sequential not parallel processes. *Arthritis Res Ther* 2016; **18**: 35.
- 53 Maksymowych WP, Wichuk S, Chiowchanwisawakit P, Lambert RG, Pedersen SJ. Fat metaplasia and backfill are key intermediaries in the development of sacroiliac joint ankylosis in patients with ankylosing spondylitis. *Arthritis Rheumatol* 2014; **66**: 2958–2967.
- 54 Kenna TJ, Brown MA. The role of IL-17-secreting mast cells in inflammatory joint disease. *Nat Rev Rheumatol* 2013; **9**: 375–379.
- 55 Smith JA. Update on ankylosing spondylitis: current concepts in pathogenesis. *Curr Allergy Asthma Rep* 2015; **15**: 489.
- 56 Pinto SM, Manda SS, Kim MS, Taylor K, Selvan LD, Balakrishnan L *et al*. Functional annotation of proteome encoded by human chromosome 22. *J Proteome Res* 2014; **13**: 2749–2760.
- 57 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.
- 58 Phipson B, Smyth GK. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol* 2010; **9** Article39.

Supplementary Information accompanies this paper on Genes and Immunity website (<http://www.nature.com/gene>)