

Published in final edited form as:

*Nat Genet.* ; 43(8): 761–767. doi:10.1038/ng.873.

## Interaction between *ERAP1* and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility

David M Evans<sup>1,43</sup>, Chris C A Spencer<sup>2,43</sup>, Jennifer J Pointon<sup>3</sup>, Zhan Su<sup>2</sup>, David Harvey<sup>3</sup>, Grazyna Kochan<sup>4</sup>, Udo Oppermann<sup>4</sup>, Alexander Dilthey<sup>1</sup>, Matti Pirinen<sup>1</sup>, Millicent A Stone<sup>5</sup>, Louise Appleton<sup>3</sup>, Loukas Moutsianas<sup>6</sup>, Stephen Leslie<sup>7</sup>, Tom Wordsworth<sup>3</sup>, Tony J Kenna<sup>8</sup>, Tugce Karaderi<sup>3</sup>, Gethin P Thomas<sup>8</sup>, Michael M Ward<sup>9</sup>, Michael H Weisman<sup>10</sup>, Claire Farrar<sup>3</sup>, Linda A Bradbury<sup>8</sup>, Patrick Danoy<sup>8</sup>, Robert D Inman<sup>11</sup>, Walter Maksymowych<sup>12</sup>, Dafna Gladman<sup>11</sup>, Proton Rahman<sup>13</sup>, Spondyloarthritis Research Consortium of Canada (SPARCC), Ann Morgan<sup>14</sup>, Helena Marzo-Ortega<sup>14</sup>, Paul Bowness<sup>3</sup>, Karl Gaffney<sup>15</sup>, J S Hill Gaston<sup>16</sup>, Malcolm Smith<sup>17</sup>, Jacome Bruges-Armas<sup>18,19</sup>, Ana-Rita Couto<sup>18</sup>, Rosa Sorrentino<sup>20</sup>, Fabiana Paladini<sup>20</sup>, Manuel A Ferreira<sup>21</sup>, Huji Xu<sup>22</sup>, Yu Liu<sup>22</sup>, Lei Jiang<sup>22</sup>, Carlos Lopez-Larrea<sup>23</sup>, Roberto Díaz-Peña<sup>23</sup>, Antonio López-Vázquez<sup>23</sup>, Tetyana Zayats<sup>1</sup>, Gavin Band<sup>2</sup>, Céline Bellenguez<sup>2</sup>, Hannah Blackburn<sup>23</sup>, Jenefer M Blackwell<sup>25,26</sup>, Elvira Bramon<sup>24</sup>, Suzannah J Bumpstead<sup>24</sup>, Juan P Casas<sup>27</sup>, Aiden Corvin<sup>28</sup>, Nicholas Craddock<sup>29</sup>, Panos Deloukas<sup>24</sup>, Serge Dronov<sup>24</sup>, Audrey Duncanson<sup>30</sup>, Sarah Edkins<sup>24</sup>, Colin Freeman<sup>2</sup>, Matthew Gillman<sup>24</sup>, Emma Gray<sup>24</sup>, Rhian Gwilliam<sup>24</sup>, Naomi Hammond<sup>24</sup>, Sarah E Hunt<sup>24</sup>, Janusz Jankowski<sup>31</sup>, Alagurevathi Jayakumar<sup>24</sup>, Cordelia Langford<sup>24</sup>, Jennifer Liddle<sup>24</sup>, Hugh S Markus<sup>32</sup>, Christopher G Mathew<sup>33</sup>, Owen T McCann<sup>24</sup>, Mark I McCarthy<sup>34</sup>, Colin N A Palmer<sup>35</sup>, Leena Peltonen<sup>24</sup>, Robert Plomin<sup>36</sup>, Simon C Potter<sup>24</sup>, Anna Rautanen<sup>24</sup>, Radhi Ravindrarajah<sup>24</sup>, Michelle Ricketts<sup>24</sup>, Nilesh Samani<sup>37</sup>, Stephen J Sawcer<sup>38</sup>, Amy Strange<sup>2</sup>, Richard C Trembath<sup>33</sup>, Ananth C Viswanathan<sup>39,40</sup>, Matthew Waller<sup>24</sup>, Paul Weston<sup>24</sup>, Pamela Whittaker<sup>24</sup>, Sara Widaa<sup>24</sup>, Nicholas W Wood<sup>41</sup>, Gilean McVean<sup>2</sup>, John D Reveille<sup>42</sup>, B Paul Wordsworth<sup>3</sup>, Matthew A Brown<sup>8,44</sup>, Peter Donnelly<sup>2,44</sup>, The Australo-Anglo-American Spondyloarthritis Consortium (TASC)<sup>1</sup>, and the Wellcome Trust Case Control Consortium 2 (WTCCC2)<sup>1</sup>

© 2011 Nature America, Inc. All rights reserved.

Correspondence should be addressed to P. Donnelly (peter.donnelly@well.ox.ac.uk) or M.A.B. (matt.brown@uq.edu.au).

<sup>43</sup>These authors contributed equally to this work.

<sup>44</sup>These authors jointly directed this work.

<sup>1</sup>A full list of members is provided in the Supplementary Note.

**URLs.** SNPTEST, <http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html>; HapMap, <http://hapmap.ncbi.nlm.nih.gov/>.

Note: Supplementary information is available on the Nature Genetics website.

**AUTHOR CONTRIBUTIONS** M.A.B., L.A.B., C. Farrar, J.D.R., J.J.P., B.P.W., D.G., W.M. and P.R. oversaw cohort collection for the discovery and replication datasets. The WTCCC2 DNA, genotyping, data quality control and informatics group (S.J.B., S.D., S.E., E.G., C.L. and L.P.) executed GWAS sample handling, genotyping and quality control. The WTCCC2 data and analysis group (A.S., C.C.A.S., G.B., C.B., C. Freeman and P. Donnelly), D.M.E. and M.A.B. performed statistical analyses. M.A.B., D.M.E., C.C.A.S. and P. Donnelly contributed to writing the manuscript. The WTCCC2 management committee (J.M.B., E.B., M.A.B., J.P.C., A.C., P. Deloukas, P. Donnelly (chairperson), A. Duncanson, J.J., J.L., H.S.M., C.G.M., C.N.A.P., L.P., R.P., A.R., S.J.S., R.C.T., A.C.V. and N.W.W.) monitored the execution of the GWAS. D.H., G.K. and U.O. performed analyses of recombinant ERAP1 function. T.J.K. and G.P.T. performed gene expression, cell count and ERAP1 sheddase functional studies. Other authors contributed variously to sample collection and all other aspects of the study. All authors reviewed the final manuscript.

**COMPETING FINANCIAL INTERESTS** The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

<sup>1</sup>Medical Research Council (MRC) Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, UK. <sup>2</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>3</sup>National Institute for Health Research Musculoskeletal Biomedical Research Unit, Nuffield Orthopaedic Centre, Headington, Oxford, UK. <sup>4</sup>Structural Genomics Consortium, University of Oxford, Oxford, UK. <sup>5</sup>University of Bath, Bath, UK. <sup>6</sup>Department of Statistics, University of Oxford, Oxford, UK. <sup>7</sup>Department of Clinical Pharmacology, University of Oxford, Oxford, UK. <sup>8</sup>University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, Australia. <sup>9</sup>National Institute of Arthritis and Musculoskeletal and Skin Diseases, US National Institutes of Health (NIH), Bethesda, Maryland, USA. <sup>10</sup>Department of Medicine/Rheumatology, Cedars-Sinai Medical Centre, Los Angeles, California, USA. <sup>11</sup>University of Toronto, Toronto, Ontario, Canada. <sup>12</sup>Department of Medicine, University of Alberta, Edmonton, Alberta, Canada. <sup>13</sup>Memorial University, St. John's, Newfoundland, Canada. <sup>14</sup>National Institute for Health Research (NIHR)–Leeds Musculoskeletal Biomedical Research Unit, University of Leeds, Leeds, UK. <sup>15</sup>Department of Rheumatology, Norfolk & Norwich University Hospital, Norfolk, UK. <sup>16</sup>Department of Medicine, University of Cambridge, Addenbrookes Hospital, Cambridge, UK. <sup>17</sup>Repatriation General Hospital, Adelaide, Australia. <sup>18</sup>Serviço Especializado De Epidemiologia E Biologia Molecular, Hospital de Santo Espírito, Angra do Heroísmo, Terceira, The Azores, Portugal. <sup>19</sup>Genetics and Arthritis Research Group, Institute for Molecular and Cell Biology (IMCB), University of Porto, Porto, Portugal. <sup>20</sup>Department of Biology and Biotechnology 'Charles Darwin', Sapienza University of Rome, Rome, Italy. <sup>21</sup>Queensland Institute of Medical Research, Brisbane, Australia. <sup>22</sup>Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, The Second Military Medical University Hospital, Shanghai, China. <sup>23</sup>Department of Immunology, Hospital Universitario Central de Asturias, Oviedo, Spain. <sup>24</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. <sup>25</sup>Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Perth, Australia. <sup>26</sup>Genetics and Infection Laboratory, Cambridge Institute of Medical Research, Addenbrooke's Hospital, Cambridge, UK. <sup>27</sup>Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. <sup>28</sup>Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland. <sup>29</sup>Department of Psychological Medicine, School of Medicine, Cardiff University, Cardiff, Wales. <sup>30</sup>Molecular and Physiological Sciences, The Wellcome Trust, London, UK. <sup>31</sup>Centre for Gastroenterology, Bart's and the London School of Medicine and Dentistry, London, UK. <sup>32</sup>Clinical Neurosciences, St. George's University of London, London, UK. <sup>33</sup>Division of Genetics and Molecular Medicine, King's College London, London, UK. <sup>34</sup>Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, Oxford, UK. <sup>35</sup>Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee, UK. <sup>36</sup>Social, Genetic and Developmental Psychiatry Centre, King's College London Institute of Psychiatry, Denmark Hill, London, UK. <sup>37</sup>Department of Cardiovascular Science, University of Leicester, Glenfield General Hospital, Leicester, UK. <sup>38</sup>University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, Cambridge, UK. <sup>39</sup>Glaucoma Research Unit, Moorfields Eye Hospital NHS Foundation Trust, London, UK. <sup>40</sup>Department of Genetics, University College London Institute of Ophthalmology, London, UK. <sup>41</sup>Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK. <sup>42</sup>Rheumatology and Clinical Immunogenetics, University of Texas Health Science Center at Houston, Houston, Texas, USA.

## Abstract

Ankylosing spondylitis is a common form of inflammatory arthritis predominantly affecting the spine and pelvis that occurs in approximately 5 out of 1,000 adults of European descent. Here we report the identification of three variants in the *RUNX3*, *LTBRTNFRSF1A* and *IL12B* regions convincingly associated with ankylosing spondylitis ( $P < 5 \times 10^{-8}$  in the combined discovery and replication datasets) and a further four loci at *PTGER4*, *TBKBPI*, *ANTXR2* and *CARD9* that

show strong association across all our datasets ( $P < 5 \times 10^{-6}$  overall, with support in each of the three datasets studied). We also show that polymorphisms of *ERAP1*, which encodes an endoplasmic reticulum aminopeptidase involved in peptide trimming before HLA class I presentation, only affect ankylosing spondylitis risk in HLA-B27-positive individuals. These findings provide strong evidence that HLA-B27 operates in ankylosing spondylitis through a mechanism involving aberrant processing of antigenic peptides.

---

Ankylosing spondylitis is the prototypic spondyloarthropathy, one of a group of conditions which also includes psoriatic arthritis, reactive arthritis and arthritis complicating inflammatory bowel disease. Ankylosing spondylitis is highly heritable (heritability >90%)<sup>1,2</sup> and familial (sibling recurrence risk ratio >52)<sup>3</sup>. Men are affected 2–3 times more frequently than women. The disease is strongly associated with HLA-B27; however, only 1–5% of HLA-B27-positive individuals develop ankylosing spondylitis, and there is strong epidemiologic evidence to suggest that other genes are involved<sup>2–5</sup>. Spondyloarthropathy is characterized histopathologically by inflammation involving the attachments of tendons and ligaments to bone (entheses). This initial enthesitis is followed by bone formation, leading to fusion (ankylosis) of the affected joints. Effective treatments are available to suppress the inflammatory component of ankylosing spondylitis, but no treatment has yet been shown to improve its long-term outcome, particularly the loss of movement caused by bony ankylosis. In addition to its strong association with HLA-B27, ankylosing spondylitis has been convincingly associated with loci containing the genes *IL23R*, *ERAP1* and *KIF21B* and with the 2p15 and 21q22 intergenic regions<sup>6–8</sup>.

In the current study, as part of the Wellcome Trust Case Control Consortium 2 (WTCCC2), we performed a genome-wide association study of 1,787 British and Australian affected individuals (cases) of European ancestry fulfilling the modified New York criteria for ankylosing spondylitis<sup>9</sup> and 4,800 common controls, following quality control steps. We genotyped cases with the Illumina 660W-Quad microarray and genotyped controls with custom Illumina Human 1.2M-Duo chips. We combined these data with the findings of non-overlapping cases and controls from the Australo-Anglo-American Spondyloarthritis Consortium (TASC)<sup>6</sup>, establishing a discovery set comprising 3,023 cases and 8,779 controls (Online Methods). We then performed a replication study in an independent cohort including 2,111 cases and 4,483 controls from Australia, Great Britain and The Spondyloarthritis Research Consortium of Canada (SPARCC).

## RESULTS

### Primary association findings

We saw genome-wide significant association ( $P < 5 \times 10^{-8}$  in the combined discovery and replication datasets) at SNPs close to or within previously established disease-associated loci, including *HLA-B*, *ERAP1*, *IL23R*, *KIF21B* and the gene deserts 2p15 and 21q22 (Table 1 and Supplementary Fig. 1), as well as at three new loci: *RUNX3*, *IL12B* and *LTBR-TNFRSF1A* (Table 2). We also found strong evidence of association for SNPs in or near *ANTXR2* ( $P = 9.4 \times 10^{-8}$ ), *CARD9* ( $P = 1.1 \times 10^{-6}$ ), *PTGER4* ( $P = 2.6 \times 10^{-7}$ ) and *TBKBPI* ( $P = 5.3 \times 10^{-8}$ ), with support in all three datasets (Table 2, Supplementary Fig. 2 and Supplementary Tables 1–3). SNPs around two of these loci, *ANTXR2* and *CARD9*, have shown strong evidence of association in previous studies<sup>6,10</sup>.

To investigate the new associations more closely, we performed further statistical analysis and functional work on *RUNX3* and *PTGER4*. *RUNX3* encodes a transcription factor critically involved in CD8 lymphocyte differentiation<sup>11</sup>. Expression of *RUNX3* is driven by IL-7 (ref. 11); in the current study, we saw moderate association with SNPs in *IL7R*

(rs13172042,  $P = 8.0 \times 10^{-5}$  and rs1993879,  $P = 9.0 \times 10^{-5}$ ). In a previous genome-wide association study of peripheral blood cell subset counts, SNPs in *RUNX3* that were associated with ankylosing spondylitis in the current study also showed association with decreased CD8 lymphocyte counts (rs10903118,  $P = 0.0037$ , determining 0.42% of CD8 count variance and rs11249215,  $P = 0.0042$ , determining 0.41% of CD8 count variance)<sup>12</sup>. Consistent with this observation, we observed that ankylosing spondylitis cases on no biological therapy have lower CD8 lymphocyte counts than age-matched controls (Supplementary Fig. 3), suggesting that the *RUNX3* association with ankylosing spondylitis may operate through an effect on CD8 lymphocyte counts.

*PTGER4* is a component of the ‘mechanostat’ anabolic bone response to physical stress, including at the site of insertion of tendons and ligaments into bone (entheses), a site characteristically involved in ankylosing spondylitis. We therefore tested expression of *PTGER4* in synovial biopsy samples from spondyloarthritis cases compared with controls (either healthy or with non-inflammatory arthritis) and found increased expression in spondyloarthritis samples (1.9-fold change,  $P = 0.04$ , two-tailed test).

### Secondary signals at disease-associated loci

To look for regions where the pattern of association is not sufficiently characterized by the most significant SNP, we ran the program GENECLUSTER on the WTCCC2 data<sup>13</sup>. Two loci, *ERAPI* and *IL23R*, showed strong evidence for a second SNP; the Bayes factor, which is the ratio of the probability of the data under a two mutation model relative to a one mutation model<sup>13,14</sup>, was 12.3 at *ERAPI* and 28.8 at *IL23R*.

At both loci, the pattern of association was consistent with a primary effect at the marginally most significant SNP (rs30187 at *ERAPI* and rs11209026 at *IL23R*) and a secondary effect at a nearby SNP (rs10050860 or rs17482078, two SNPs with  $r^2 = 1$ , at *ERAPI* and rs11209032 at *IL23R*). Frequentist analysis conditioning on the primary SNP confirmed the secondary SNP as explaining a significant additional fraction of risk at both loci in both the discovery and replication data (*IL23R*, rs11209032, WTCCC2  $P = 7.9 \times 10^{-6}$ , TASC  $P = 0.0018$ , replication  $P = 0.0018$ ; *ERAPI*, rs10050860, WTCCC2  $P = 0.0025$ , TASC  $P = 0.026$ , replication  $P = 0.00023$ ; Table 3 and Supplementary Fig. 4). Around *IL23R*, the minor allele at rs11209026 (which is inferred to be non-ancestral) is strongly protective, whereas the minor allele of rs11209032 increases disease risk. In *ERAPI*, the major allele of rs30187 and the minor allele of rs10050860 are protective (and both are inferred to be non-ancestral). In combination, the protective effect was particularly strong; HLA-B27-positive individuals homozygous for protective variants at rs30187 and rs10050860 had approximately 3–4 times lower disease risk than HLA-B27-positive individuals with risk predisposing alleles at these loci (Supplementary Fig. 4).

### Major histocompatibility complex (MHC) studies

In addition to our association analyses, we tested genotyped and imputed SNPs across the MHC for their ability to tag HLA-B27. We identified one SNP, rs4349859, that strongly tagged *HLA-B27* and had high sensitivity (98.0%) and specificity (99.0%) in 538 cases and 741 controls of European origin from Australia and Great Britain (Supplementary Table 4). This SNP also showed the strongest association with ankylosing spondylitis in our study. A second SNP, rs13202464, showed slightly higher sensitivity (98.7%) but slightly lower specificity (98.0%) for *HLA-B27* than rs4349859. Further studies will be required to determine the value of typing both SNPs over and above rs4349859 alone to impute *HLA-B27* genotype. Controlling for rs4349859 and rs13202464, there were some residual signals in the MHC, which could warrant further investigation, but no other SNP across the MHC achieved  $P < 4.0 \times 10^{-6}$  in the combined discovery and replication datasets.

rs4349859 lies 41-kb centromeric of *HLA-B* and 5.4-kb telomeric of *MICA*, which neighbors *HLA-B*. *MICA* has previously been associated with ankylosing spondylitis<sup>15,16</sup>; in order to distinguish whether our finding was caused by linkage disequilibrium with HLA-B27 or whether it indicated a potential disease-causative variant itself, we tested the ability of rs4349859 to tag HLA-B27 subtypes with differential association with ankylosing spondylitis. These studies confirmed that rs4349859 tags all the major European-ancestry ankylosing spondylitis-associated subtypes, HLA-B\*2702, HLA-B\*2705 and HLA-B\*2708 (ref. 17), but also tags HLA-B\*2709, which is not associated with ankylosing spondylitis<sup>18</sup>. However, rs4349859 does not effectively tag the African ankylosing spondylitis-associated subtype HLA-B\*2703 nor the Asian HLA-B27 subtypes HLA-B\*2704, HLA-B\*2706 or HLA-B\*2707 (ref. 19), although it does tag HLA-B\*2705 in Asian carriers (Supplementary Table 4). Further, neither rs4349859 nor rs1320464 was more sensitive or specific than HLA-B27 for ankylosing spondylitis in cases and controls typed for each polymorphism in the WTCCC2-TASC dataset (data not shown).

### Gene-gene interaction studies

We tested loci known to be associated with ankylosing spondylitis, including those identified in the current study, for gene-gene interaction (defined as a departure from additivity on the log-odds scale). We observed strong evidence for gene-gene interaction between HLA-B27 (as tagged by rs4349859, with a dominant model<sup>20</sup>) and the nonsynonymous SNP (nsSNP) rs30187 in *ERAP1* in the WTCCC2, TASC and replication datasets (WTCCC2  $P = 0.02$ , TASC  $P = 0.014$ , replication  $P = 0.0019$ , combined  $P = 7.3 \times 10^{-6}$ ; see also Supplementary Table 5). The interaction takes a particularly simple form: in HLA-B27-positive cases and controls, *ERAP1* was associated with ankylosing spondylitis, whereas we saw no association with *ERAP1* in HLA-B27-negative cases (Fig. 1). In contrast, SNPs at other risk loci, including *IL23R*, *KIF21B*, *IL12B* and the intergenic regions 2p15 and 21q22, were associated with ankylosing spondylitis in both HLA-B27-positive and HLA-B27-negative disease (Supplementary Table 6).

Odds ratios and 95% confidence intervals for each of the possible genotype combinations are presented in Supplementary Table 7. We tested the robustness of the HLA-B27-*ERAP1* interaction by also fitting the full (but less powerful) 4-degree-of-freedom interaction test (Online Methods). We observed evidence for interaction between rs4349859 and rs30187 in the WTCCC2 (WTCCC2  $\chi^2 = 9.77$ ,  $P = 0.04$ ) and replication cohorts (replication  $\chi^2 = 10.05$  (4 degrees of freedom),  $P = 0.04$ ), whereas the pattern of odds ratios in the TASC dataset indicated a trend in the same direction (TASC  $\chi^2 = 7.03$ ,  $P = 0.13$ ). Combining the evidence across cohorts provided strong evidence for the existence of an interaction (combined  $\chi^2 = 23.28$ ,  $P = 0.0001$ ), showing that our results are robust and are not likely to be an artifact.

### Investigation of ERAP1 function

ERAP1 has two reported functions. First, it may act as a ‘molecular ruler’ within the endoplasmic reticulum, trimming peptides that have been partially processed by the proteasome (typically 9–17 amino acids in length) down to nine amino acids in length, the optimal length for HLA class I binding and presentation<sup>21,22</sup>. Second, *in vitro* studies have suggested that ERAP1 may function as a ‘shedase’, cleaving cytokine receptors off the cell surface, including IL-6R, IL-1R2 and TNFR<sup>23–25</sup>. We tested the ability of ERAP1 to cleave TNFR and IL-6R from cell surfaces by measuring the appearance of these receptors in cell culture supernatants from single-cell suspensions prepared from *Erap1*<sup>-/-</sup> and C57BL/6 control mouse spleens (Supplementary Fig. 5). We observed no difference in the levels of these receptors over time, indicating that ERAP1 does not have a major influence on cytokine receptor trimming, at least in mice.



In a preliminary *in vitro* examination of the effects of allelic variants on the function of ERAP1 in peptide trimming, we synthesized recombinant ERAP1 with the wild-type sequence or the ankylosing spondylitis-associated variants p.Lys528Arg (rs30187), p.Arg725Gln (rs17482078) or p.Asp575Asn (rs10050860) and tested their ability to trim an ERAP1 substrate<sup>26</sup>. In the discovery dataset, the second SNP at *ERAP1*, rs10050860, was in complete linkage disequilibrium with rs17482078. Both the protective rs30187 and rs17482078 alleles had ~40% slower rates of substrate trimming than wild-type ERAP1 ( $P=0.050$ , Wilcoxon rank sum test, one-tailed exact test), suggesting that both represent loss-of-function variants (Fig. 2); the proteolytic rate of rs10050860 was similar to wild-type ERAP1.

## DISCUSSION

This study increases the number of genetic loci convincingly associated with ankylosing spondylitis to nine (*IL23R*, *RUNX3*, *KIF21B*, 2p15, *IL12B*, *ERAP1*, *HLA-B*, *LTBR-TNFRSF1A* and 21q22), identifies a further four loci that show strong evidence of association and are likely to be involved in disease pathogenesis (*ANTXR2*, *PTGER4*, *CARD9* and *TBKBPI*) and greatly expands our understanding of the etiopathogenesis of the condition, including the likely mechanism by which HLA-B27 operates in the disease. These new genetic findings further implicate the involvement of the IL-23–IL-23R–IL-17 pathway in ankylosing spondylitis. Although further studies will be required to identify the true disease-associated variants, each associated locus contains candidate genes of strong biological interest in ankylosing spondylitis, which are discussed below. *IL12B* encodes IL12p40, which is a shared component of both IL-12 and IL-23. Association has previously been reported between SNPs in *IL23R* and ankylosing spondylitis<sup>7</sup>. SNPs in or near *CARD9* and *PTGER4* (encoding prostaglandin E receptor 4, EP4 subtype) have also been associated with Crohn's disease previously<sup>27,28</sup>, and variants within *IL12B* and *IL23R* have been associated with both Crohn's disease and psoriasis<sup>29–31</sup>. PGE<sub>2</sub> acts through *PTGER4* to induce the production of IL-23 and IL-17 and to promote expansion of T<sub>H</sub>17 lymphocyte counts; T<sub>H</sub>17 lymphocyte counts and IL-17 levels are both known to be elevated in ankylosing spondylitis<sup>32</sup>. Consistent with this, we observed upregulation of *PTGER4* expression in spondyloarthritis synovium. *CARD9* mediates signals from dectin-1 and -2, which are innate immunity receptors for β-glucan, a component of fungal and some bacterial cell walls. β-glucan induces PGE<sub>2</sub> production, which in turn promotes the expansion of T<sub>H</sub>17 cells<sup>33</sup>. We and others have recently reported that SKG mice treated with β-glucan develop spondyloarthritis and Crohn's disease<sup>34</sup>; this mouse model is characterized by T<sub>H</sub>17 activation<sup>35</sup>. These findings are thus consistent with a model of human ankylosing spondylitis in which ubiquitous pathogens carrying β-glucan induce disease by stimulating the IL-23–IL-23R–IL-17 pathway.

Despite considerable research since the discovery of the association of HLA-B27 with ankylosing spondylitis in the early 1970s, the mechanism explaining the involvement of HLA-B27 with the disease is unclear. Current theories can essentially be divided into those that favor canonical functions of HLA-B27 in peptide presentation (the 'arthritogenic peptide' theory) and those that invoke aberrant properties of HLA-B27 itself, including the propensity to form HLA-B27 homodimers and/or induce endoplasmic reticulum stress caused by slow and inefficient folding of nascent HLA-B27 molecules. A further possible hypothesis is that HLA-B27 acts as a tag for another, nearby, true susceptibility allele. Our findings that no other MHC SNP was more strongly associated with ankylosing spondylitis than rs4349859 or HLA-B27 indicate that this 'linked gene' theory is unlikely. The finding that rs4349859 does not tag HLA-B27 subtypes common in Asian populations but does tag HLA-B\*2709, a European HLA-B27 subtype not associated with ankylosing spondylitis, indicates that rs4349859 is not disease causative in its own right. However, it does tag the

ancestral HLA-B\*2705 subtype, and other European HLA-B27 subtypes, with high sensitivity and specificity within the level of accuracy of HLA-B27 typing itself but at far lower cost and difficulty.

The interaction between HLA-B27 and *ERAP1* is also highly informative regarding potential mechanisms of disease in ankylosing spondylitis. The restriction of the association of *ERAP1* with ankylosing spondylitis to HLA-B27-positive cases is consistent with disease models in which aberrant trimming of peptides or presentation by ERAP1 and HLA-B27 are involved in the pathogenesis of HLA-B27-associated disease. HLA-B27- and ERAP1-negative disease is unlikely to be caused by a similar mechanism, and the fact that overexpression of TNF alone is sufficient to cause spondyloarthritis in mice<sup>36</sup> suggests that overexpression or signaling by pro-inflammatory cytokines alone may be sufficient to cause ankylosing spondylitis. Although the findings of our studies of ERAP1 function using recombinant ERAP1 protein require much further experimental validation, they suggest that *ERAP1* influences disease risk in ankylosing spondylitis, with associated, non-ancestral alleles reducing the disease risk by a mechanism involving reduced peptide trimming and, likely as a consequence, altered peptide presentation by HLA-B27. This finding, and the strong protective effect of *ERAP1* variants in ankylosing spondylitis, suggest that inhibition of ERAP1 may be effective in ankylosing spondylitis, at least in HLA-B27-positive disease.

Modeling studies in ankylosing spondylitis<sup>3</sup> and in many other common diseases<sup>37,38</sup>, as well as findings in many plant and animal species, suggest that the best genetic models for common diseases involve gene-gene interaction. Very few well-replicated examples of gene-gene interaction in common polygenic diseases have been reported to date, making the finding of interaction between HLA-B27 and *ERAP1* in ankylosing spondylitis of particular interest. A similar observation was made in the overlapping WTCCC2 study of psoriasis<sup>39</sup>, with gene-gene interaction observed between *ERAP1* and HLA-Cw6.

We present here the first convincing evidence of associations in HLA-B27-negative ankylosing spondylitis. HLA-B27-negative ankylosing spondylitis has very similar clinical presentation and disease course as HLA-B27-positive disease, with a similar natural history but with a later age of onset and lower incidence of uveitis<sup>40,41</sup>. Few families have been reported with recurrence of ankylosing spondylitis in the absence of HLA-B27, no studies have examined the heritability of ankylosing spondylitis in its absence and no loci have previously been convincingly associated with ankylosing spondylitis in HLA-B27-negative cases, probably because only small studies have been performed addressing this. Our findings indicate both substantial overlap and differences in genetic susceptibility between HLA-B27-positive and -negative ankylosing spondylitis and suggest a shared etiopathogenesis.

This study provides further evidence of the involvement of TNF pathways in ankylosing spondylitis pathogenesis. TNF antagonists are highly effective in suppressing inflammation in ankylosing spondylitis, and mice overexpressing TNF develop spondyloarthritis<sup>36</sup>. In the current study, we observed association with SNPs at chromosome 12p13 between *LTBR* (lymphotoxin beta receptor) and *TNFRSF1A* (tumor necrosis factor receptor 1 (TNFR1)) (rs11616188, WTCCC2  $P = 1.2 \times 10^{-5}$ , TASC  $P = 6.2 \times 10^{-8}$ , replication  $P = 0.0036$ , combined  $P = 4.1 \times 10^{-12}$ ) (Supplementary Fig. 2). A mouse model with constitutive overexpression of TNF that develops inflammatory bowel disease and spondyloarthritis has been reported, in which the arthritis has been shown to be dependent on TNFR1 expression and in which TNFR1 expression in mesenchymal tissue alone is sufficient to permit disease<sup>36</sup>. This is consistent with a role for TNFR1 in ankylosing spondylitis, but further studies will be required to determine if the genetic polymorphisms at this locus are associated with ankylosing spondylitis because of effects on *LTBR*, *TNFRSF1A* or both.

We also observed association at chromosome 17q21 near *TBKBPI* (encoding TBK binding protein 1), a component of the TNF receptor signaling pathway (rs8070463, WTCCC2  $P=0.0025$ , TASC  $P=1.6 \times 10^{-5}$ , replication  $P=0.015$ , combined  $P=5.3 \times 10^{-8}$ ) (Supplementary Fig. 2). *TRADD*, which has previously shown suggestive association<sup>6,42</sup> and which also shows suggestive association with ankylosing spondylitis in this study encodes the TNF-receptor-associated death domain protein, also part of the TNF signaling pathway (rs9033, WTCCC2  $P=9.0 \times 10^{-6}$ , TASC  $P=0.32$ , combined  $P=4.9 \times 10^{-5}$ ). Further research will be required to determine the precise mechanism underlying these associations.

The association of SNPs in and around *RUNX3* provides further support for the hypothesis that HLA-B27 is involved in ankylosing spondylitis by a mechanism involving peptide presentation. The canonical function of HLA class I molecules is to present antigen to CD8 lymphocytes. *RUNX3* is a key gene involved in CD8 lymphocyte differentiation, and we have shown here that ankylosing spondylitis cases have reduced CD8 lymphocyte counts and that the *RUNX3* SNPs associated with ankylosing spondylitis are also associated with reduced CD8 lymphocyte counts in healthy individuals. These findings are consistent with a key role for CD8 lymphocytes in ankylosing spondylitis pathogenesis, likely in response to antigenic peptides presented by HLA-B27.

In conclusion, this study has identified further non-MHC genetic loci associated with ankylosing spondylitis. Our data strongly imply that HLA-B27 itself, rather than a nearby linked gene, is directly involved in ankylosing spondylitis susceptibility. We have also shown that HLA-B27-positive and -negative ankylosing spondylitis cases differ in association with *ERAPI1*, which is involved in endoplasmic reticulum peptide processing before HLA class I assembly and peptide presentation. This represents one of the first convincingly replicated examples of gene-gene interaction in any common disease and indicates that the likely mechanism by which HLA-B27 induces ankylosing spondylitis involves aberrant presentation or handling of peptides.

## ONLINE METHODS

### Samples

For cases, ankylosing spondylitis was defined according to modified New York criteria<sup>9</sup>. A total of 1,787 DNA samples from unrelated individuals of European ancestry were included in the GWAS discovery set, having passed genotyping control filters (see below and Supplementary Table 8). The replication set was derived from individuals with ankylosing spondylitis recruited from the UK, Australia and Canada; 2,011 DNA samples passed quality control for replication. All case participants gave written, informed consent, and the study was approved by the relevant research ethics authorities at each participating center.

For controls, a total of 4,800 individuals from the WTCCC2 common control set passed the quality control filters (see below) and were used in the discovery GWAS. This included 2,485 healthy blood donors from the UK Blood Service (UKBS) collection and 2,315 individuals from the 1958 Birth Cohort dataset (see below and Supplementary Table 8). The number of replication control samples passing quality control totaled 4,483, comprising 2,831 from the People of the British Isles collection and 1,652 Canadian controls.

### GWAS genotyping

Following careful DNA quality control, including gender verification by genotyping in comparison with patient data manifests, samples were genotyped at the Wellcome Trust Sanger Institute on the Illumina Infinium platform. Samples from the case collection were genotyped on the Human660W-Quad (a custom chip designed by WTCCC2 and comprising



Human550 and a set of ~6,000 common copy number variants from the Structural Variation Consortium<sup>43</sup>). Samples from the control collections were genotyped on the Human1.2M-Duo (a WTCCC2 custom array comprising Human1M-Duo and the common copy number variant content described above). Bead intensity data was processed and normalized for each sample in BeadStudio; data for successfully genotyped samples was extracted and genotypes were called within collections using Illuminus<sup>44</sup>.

### GWAS quality control

We identified and removed samples whose genome-wide patterns of diversity differed from those of the collection at large, interpreting them as likely to be caused by biases or artifacts. To do so, we used a Bayesian clustering approach to infer outlying individuals on the basis of call rate, heterozygosity, ancestry and average probe intensity. To obtain a set of putatively unrelated individuals, we used a Hidden Markov model to infer identity by descent along the genome and then iteratively removed individuals to obtain a set with pairwise identity by descent less than 5%. These methods are described elsewhere<sup>39,45</sup>. Sample exclusions before imputation are outlined in Supplementary Table 8.

For SNPs, a measure of (Fisher) information for the allele frequency at each SNP was calculated using SNPTEST (see URLs). Autosomal SNPs were excluded if this information measure was below 0.975, if the minor allele frequency was less than 0.1%, if the SNP had >2% missing data or if the Hardy-Weinberg  $P$  value was lower than  $5 \times 10^{-20}$  in the case data or in either of the control collections.

### Imputation

For WTCCC2 data, genotypes were imputed using IMPUTE2 (ref. 46), which adopts a two-stage approach using both haploid and diploid reference panels. For the haploid reference panel, we used HapMap2 and HapMap3 SNP data for the 120 non-related CEU trios (see URLs). For the diploid reference panel, we used 1958 Birth Cohorts and UK Blood Service control data, merging genotypes from the Illumina 1.2M Duo chip and the Affymetrix Genome Wide Human SNP array 6.0. Prior to analysis with IMPUTE2, we applied standard quality control filters akin to those described above. To further protect against potential errors misleading the imputation and phasing, we checked that each genotype conformed to local patterns of linkage disequilibrium in HapMap by using a leave-one-out imputation strategy. Specifically, we ran IMPUTE<sup>46</sup> on each of the study samples in turn, both cases and controls, re-imputing known genotypes. Control individuals for which the imputed genotypes were more than 4.5% discordant with the original genotype were removed. The same rule was applied to case individuals with a discordance threshold of 6%. SNPs for which IMPUTE was confident of the imputation call but for which the genotyped data was discordant (and therefore indicative of genotyping error) were also removed if the difference between the measure of information and error rate was greater than 0.05. The final dataset comprised 1,787 cases and 4,800 controls.

After merging the imputed and genotyped data, we used the program SNPTEST to perform frequentist tests of association at each SNP under a missing data logistic regression model. Unless otherwise stated, we assumed a multiplicative model for allelic risk by encoding the genotypes at each SNP as a discrete explanatory variable with an indicator of case status as the binary response.  $P$  values were calculated using the score test implementation in SNPTEST. SNPs for which the (Fisher) information measure, as calculated by SNPTEST, was  $<0.85$  in the case collection or  $<0.95$  in either of the control groups were removed from the analysis. We also removed SNPs for which the equivalent observed (Fisher) information about the parameter describing the risk associated with the individual's genotype was  $<0.85$ . Both of these quality control steps exclude SNPs with high missingness rates or uncertain

genotypes. Only SNPs with a minor allele frequency >1% were used for subsequent analyses. Evidence for population stratification was minimal in the WTCCC2 discovery set ( $\lambda = 1.056$ ), so covariates reflecting ancestry were not included in association models (Supplementary Table 9 and Supplementary Fig. 6).

For TASC data, findings from the discovery GWAS were combined with previously reported results from 1,236 non-overlapping cases and 3,979 controls from TASC<sup>6</sup>. Data from the TASC discovery sets were imputed with Markov Chain Haplotyping software (MACH) using phased data from CEU individuals from release 22 of the HapMap project as the reference set of haplotypes<sup>47</sup>. Only SNPs common to cases and controls were used for imputation. Only SNPs that could be imputed with relatively high confidence ( $R^2 > 0.3$ ) and had a minor allele frequency >1% were used in subsequent analyses. Association analysis of imputed SNPs was performed assuming an underlying additive model with the use of the software package MACH2DAT, which regresses affection status against expected dosage at each imputed locus<sup>47</sup>. Analysis of genotyped SNPs was performed using logistic regression in PLINK<sup>48</sup>. As the evidence of population stratification was appreciable in the TASC discovery set, the first four components from an EIGENSTRAT principal components analysis were included as covariates in the model, as has been described previously<sup>6</sup>. After correction, the TASC data had an inflation factor of  $\lambda = 1.025$  (Supplementary Table 9 and Supplementary Fig. 6).

### Meta-analysis

Prior to analysis, further checks were made to exclude any cryptic sample duplication within, or between, TASC and WTCCC2 cases and controls, leading to the exclusion of a further 13 samples. In total, the sample and SNP exclusions described above resulted in the association analysis of 3,023 cases and 8,779 controls at 2,223,620 autosomal SNPs. To combine the evidence of association across the TASC and WTCCC2 data, we conducted a fixed effect meta-analysis at SNPs present in both datasets. To do so, we averaged the estimated effect size parameter associated with genotype risk across the two datasets, weighted by the inverse of the variance in the estimates. *P* values were calculated assuming the square of the combined data z-score was distributed as  $\chi^2$  with 1 degree of freedom. As an additional quality control step, we compared the estimated effect size in each dataset to the meta-analysis estimate using Cochran's *Q*, excluding SNPs with heterogeneity  $P < 0.01$ . These analyses were conducted using the statistical package R<sup>49</sup>.

### Replication genotyping and quality control

Genotyping was carried out at the Wellcome Trust Sanger Institute using the Sequenom iPLEX Gold assay. Individual samples were excluded from analysis if they had call rates <90% or if the reported gender was discordant with sex-specific markers. We removed pairs of samples showing concordance indicative of being duplicates with individuals in the replication and/or discovery sets.

### Replication and combined analysis

Association in the replication cohort was performed using logistic regression assuming an additive model and including a dichotomous covariate reflecting Canadian ancestry. Analyses were also performed in the statistical package R. We report replication *P* values for the likelihood ratio test comparing the models with and without a SNP effect on risk, which is a two-tailed test on the direction of the effect.

## Interaction, conditional and haplotype analyses

Results from the WTCCC2-TASC meta-analysis and the replication data were combined using the fixed effect model described above. Frequentist conditional analysis was performed by including the primary SNP as a covariate in the logistic regression. We then tested whether the addition of a secondary SNP better explained risk at the locus using a likelihood ratio test with 1 degree of freedom. We took the same approach to testing for interaction (defined as a departure from additivity on the log-odds scale) by comparing a model with an additive term for each variant to one which also included a multiplicative interaction term:  $y = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 * x_2$ , where  $y$  is the log odds of disease,  $\mu$  is an intercept,  $x_1$  and  $x_2$  are dummy variables coded so as to reflect underlying additive effects (0, 1 or 2) at the first and second SNPs and  $\beta_1$ ,  $\beta_2$  and  $\beta_{12}$  are regression coefficients estimated from the data. In the case of rs4349859 (which tags HLA-B27), we fitted a dominant term for  $x_1$  rather than an additive term ( $x_1 = 0$  if the genotype at rs4349859 is GG, and  $x_1 = 1$  if the genotype at rs4349859 is AG or AA). Twice the difference in log likelihood between the full and base models is distributed as a  $\chi^2$  statistic with 1 degree of freedom.

Given the large difference in magnitude between main effects at rs4349859 and the other loci, we also tested the robustness of our findings by fitting a full two-locus model with additive and dominance terms and interactions between them:  $y = \mu + \beta_{a1} x_1 + \beta_{d1} x_1 + \beta_{a2} x_2 + \beta_{d2} x_2 + \beta_{aa} x_1 * x_2 + \beta_{ad} x_1 * z_2 + \beta_{da} z_1 * x_2 + \beta_{dd} z_1 * z_2$ , where  $y$  is the log odds of disease,  $\mu$  is an intercept,  $x_1$  and  $x_2$  are dummy variables coded so as to reflect underlying additive effects (0, 1 or 2) at the first and second SNPs,  $z_1$  and  $z_2$  are dummy variables coded so as to pick up any dominance effects at the loci (0, 1 or 0), and  $\beta_{a1}$ ,  $\beta_{d1}$ ,  $\beta_{a2}$ ,  $\beta_{d2}$ ,  $\beta_{aa}$ ,  $\beta_{ad}$ ,  $\beta_{da}$  and  $\beta_{dd}$  are regression coefficients estimated from the data. Covariates were included in the model where appropriate to account for cohort and ancestry effects. The test for interaction was performed by comparing this full model to a model where only the main effects were fitted, with twice the difference in log likelihood between the models being distributed as a  $\chi^2$  statistic with 4 degrees of freedom.

As the pairs of SNPs identified at *HLA-B*, *ERAP1* and *IL23R* showed very little evidence of recombination ( $D' > 0.98$ ), we phased individual's genotypes by parsimony, excluding the small number of samples (<20 at each SNP) with genotypes consistent with rare recombination, genotyping error or repeat mutation. Haplotype risks were estimated from the replication data using a logistic regression model by including an additive term for each of the haplotypes.

The program GENECLUSTER was used to look in the WTCCC2 data for regions where the pattern of association is not sufficiently characterized by the most significant SNP<sup>13</sup>.

## Synovial tissue gene expression studies

We obtained 15 synovial biopsy tissue samples from the Synovial Tissue Bank at the Repatriation General Hospital in Adelaide, South Australia with the appropriate ethical approvals. The samples consisted of six seronegative spondyloarthritis, two ankylosing spondylitis, three osteoarthritis and four normal control biopsies. RNA was extracted from the formaldehyde-fixed paraffin-embedded tissue blocks using the Arcturus Paradise Plus Reagent System (Molecular Devices) according to the manufacturer's protocol. We used 200 ng of RNA in the Illumina Whole-Genome DASL (complementary DNA-mediated annealing, selection, extension and ligation) Gene Expression Assay. Data was processed in GenomeStudio (Illumina) and analyzed using Lumi<sup>50</sup> and BRB ArrayTools<sup>51</sup>. For analysis, ankylosing spondylitis and spondyloarthritis samples were grouped together and compared with a control group consisting of osteoarthritis (a non-inflammatory arthritis) and

unaffected samples. Unsupervised clustering showed no differences between ankylosing spondylitis and spondyloarthritis or between osteoarthritis and normal samples, supporting the grouping of these samples for analysis. However, unsupervised clustering between the ankylosing spondylitis and spondyloarthritis and osteoarthritis and normal groups showed good delineation of the two groups, with only one sample from each group being misclassified. Differentially expressed genes were identified by unpaired *t*-tests with multivariate permutation correction.

### Cytokine receptor cleavage studies

Single-cell suspensions were prepared from spleens of *Erap1*<sup>-/-</sup> and C57BL/6 control mice and plated in 48-well plates at  $5 \times 10^6$  cells per well. Cells were stimulated with plate-bound anti-CD3 (10  $\mu$ g/ml, Clone 145-2C11) and phorbol 12-myristate 13-acetate (PMA) (10 ng/ml) or left untreated. Levels of soluble TNF-R and IL-6R were quantified from cell culture supernatants obtained 24 h after stimulation TNF-R, and IL-6R levels were evaluated by ELISA according to the manufacturer's protocol (R&D Systems).

### ERAP1 peptidase activity studies

Full-length wild-type *ERAP1* was cloned into pFastBac as a C-terminal His10-tagged fusion protein using standard cloning techniques. ERAP1 substitution mutants p.Lys528Arg and p.Arg725Gln were generated by site-directed mutagenesis. Recombinant proteins were purified to homogeneity on affinity Ni-NTA, and purity was assessed by SDS-PAGE followed by Coomassie staining. Enzymatic activities were assessed as previously published<sup>26</sup>.

### CD8 lymphocyte counts in ankylosing spondylitis cases and controls

Peripheral blood was obtained from 16 individuals with active ankylosing spondylitis (erythrocyte sedimentation rate > 25 and C-reactive protein > 10) and 16 age-matched healthy controls. Peripheral blood mononuclear cells (PBMCs) were extracted using a standard density gradient centrifugation over Ficoll-Paque Plus (GE Healthcare). Extracted PBMCs were frozen in 90% FBS/10% DMSO until needed. Frozen PBMCs were thawed into RPMI + 20% FBS and washed once in RPMI + 10% FBS. PBMCs were rested in RPMI + 10% FBS for approximately 1 h at 37 °C, 5% CO<sub>2</sub> before further use. PBMCs were stained with CD3 ECD (UCHT1, Beckman Coulter) and CD8 APC AlexaFluor700 (B9.11, Beckman Coulter). Dead cells were excluded using a Live/Dead Fixable Dead Cell Stain kit (Invitrogen). Cells were acquired on a Gallios 10-color flow cytometer (Beckman Coulter) and staining was analyzed using Kaluza software (Beckman Coulter).

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We would like to thank all participating subjects with ankylosing spondylitis and healthy individuals who provided the DNA and clinical information necessary for this study. The Wellcome Trust Case Control Consortium 2 project is funded by the Wellcome Trust (083948/Z/07/Z). We also thank S. Bertrand, J. Bryant, S.L. Clark, J.S. Conquer, T. Dibling, J.C. Eldred, S. Gamble, C. Hind, A. Wilk, C.R. Stribling and S. Taylor of the Wellcome Trust Sanger Institute's Sample and Genotyping Facilities for technical assistance. The TASC study was funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) grants P01-052915 and R01-AR046208. Funding was also received from the University of Texas at Houston Clinical and Translational Science Awards grant UL1RR024188, Cedars-Sinai General Clinical Research Centre grant MO1-RR00425, Intramural Research Program, NIAMS/US National Institutes of Health and Rebecca Cooper Foundation (Australia). This study was funded, in part, by Arthritis Research UK (Grants 19536 and 18797), by the Wellcome Trust (grant number 076113) and by the Oxford Comprehensive Biomedical Research Centre ankylosing spondylitis chronic disease

cohort (theme code: A91202). The Spondyloarthritis Research Consortium of Canada (SPARCC) was funded by a National Research Initiative Award from the Arthritis Society (Canada). G.P.T. was funded by a Lions Medical Research Foundation fellowship. M.A.B. is funded by a National Health and Medical Research Council (Australia) Principal Research Fellowship, and support for this study was received from a National Health and Medical Research Council (Australia) program grant (566938) and project grant (569829) and from the Australian Cancer Research Foundation and Rebecca Cooper Medical Research Foundation. P. Donnelly was supported in part by a Wolfson-Royal Society Merit Award. We are also very grateful for the invaluable support received from the National Ankylosing Spondylitis Society (UK) and Spondyloarthritis Association of America in case recruitment. Additional financial and technical support for subject recruitment was provided by the National Institute for Health Research (NIHR), Oxford Musculoskeletal Biomedical Research Unit and NIHR Thames Valley Comprehensive Local Research Network. We acknowledge use of the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02, and we thank W. Bodmer and B. Winney for use of the People of the British Isles DNA collection, which was funded by the Wellcome Trust. We would like to thank A. Mathieu (Cagliari University) and S. Brown (Cedars-Sinai Hospital) for providing samples. The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Karolinska Institutet, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck & Co., Inc., the Novartis Research Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research and the Wellcome Trust.

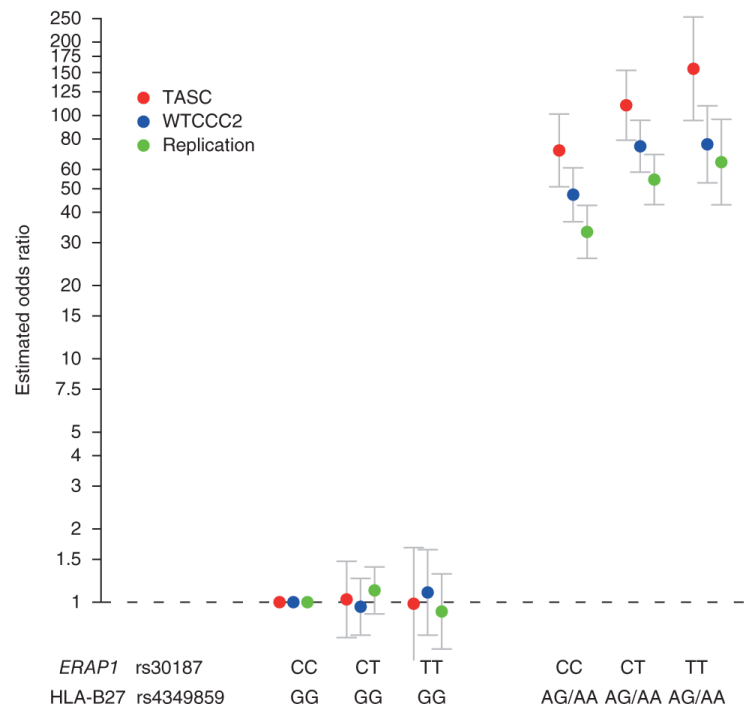
## References

1. Pedersen OB, et al. Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the relative importance of genetic vs. environmental effectors in disease causation. *Scand. J. Rheumatol.* 2008; 37:120–126. [PubMed: 18415769]
2. Brown MA, et al. Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. *Arthritis Rheum.* 1997; 40:1823–1828. [PubMed: 9336417]
3. 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. [PubMed: 11053066]
4. Calin A, Marder A, Becks E, Burns T. Genetic differences between B27 positive patients with ankylosing spondylitis and B27 positive healthy controls. *Arthritis Rheum.* 1983; 26:1460–1464. [PubMed: 6606431]
5. van der Linden S, Valkenburg H, Cats A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals: a family and population study. *Br. J. Rheumatol.* 1983; 22:18–19. [PubMed: 6606472]
6. Reveille JD, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat. Genet.* 2010; 42:123–127. [PubMed: 20062062]
7. Wellcome Trust Case Control Consortium. et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* 2007; 39:1329–1337. [PubMed: 17952073]
8. Danoy P, et al. Association of variants at 1q32 and *STAT3* with ankylosing spondylitis suggests genetic overlap with Crohn's disease. *PLoS Genet.* 2010; 6:e1001195. [PubMed: 21152001]
9. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum.* 1984; 27:361–368. [PubMed: 6231933]
10. Pointon JJ, et al. Elucidating the chromosome 9 association with AS; *CARD9* is a candidate gene. *Genes and Immunity.* 2011; 12:319–320.
11. Park JH, et al. Signaling by intrathymic cytokines, not T cell antigen receptors, specifies CD8 lineage choice and promotes the differentiation of cytotoxic-lineage T cells. *Nat. Immunol.* 2010; 11:257–264. [PubMed: 20118929]
12. Ferreira MA, et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.* 2010; 86:88–92. [PubMed: 20045101]
13. Su Z, et al. Bayesian method for detecting and characterizing allelic heterogeneity and boosting signals in genome-wide association studies. *Stat. Sci.* 2009; 24:430–450.
14. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.* 2009; 10:681–690. [PubMed: 19763151]



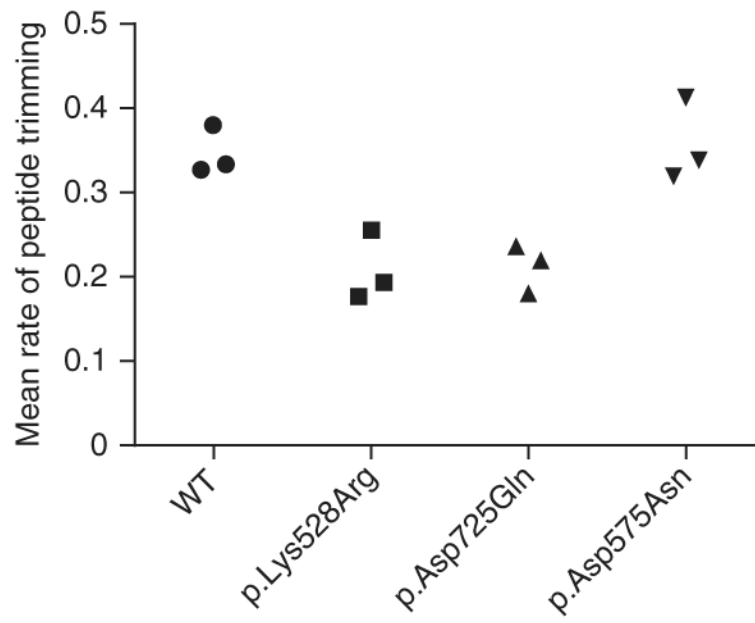
15. Amroun H, et al. Early-onset ankylosing spondylitis is associated with a functional MICA polymorphism. *Hum. Immunol.* 2005; 66:1057–1061. [PubMed: 16386647]
16. Yabuki K, et al. Triplet repeat polymorphism in the *MICA* gene in HLA-B27 positive and negative Caucasian patients with ankylosing spondylitis. *Hum. Immunol.* 1999; 60:83–86. [PubMed: 9952031]
17. Brown MA, et al. HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. *Ann. Rheum. Dis.* 1996; 55:268–270. [PubMed: 8733445]
18. D'Amato M, et al. Relevance of residue 116 of HLA-B27 in determining susceptibility to ankylosing spondylitis. *Eur. J. Immunol.* 1995; 25:3199–3201. [PubMed: 7489765]
19. López-Larrea C, et al. HLA-B27 subtypes in Asian patients with ankylosing spondylitis. Evidence for new associations. *Tissue Antigens.* 1995; 45:169–176. [PubMed: 7761976]
20. Brown MA, et al. HLA Class I and II associations of ankylosing spondylitis. *Arthritis Rheum.* 2009; 60:S437.
21. Saric T, et al. An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat. Immunol.* 2002; 3:1169–1176. [PubMed: 12436109]
22. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism. *Proc. Natl. Acad. Sci. USA.* 2005; 102:17107–17112. [PubMed: 16286653]
23. Cui X, Rouhani FN, Hawari F, Levine SJ. Shedding of the type II IL-1 decoy receptor requires a multifunctional aminopeptidase, aminopeptidase regulator of TNF receptor type 1 shedding. *J. Immunol.* 2003; 171:6814–6819. [PubMed: 14662887]
24. Cui X, Rouhani FN, Hawari F, Levine SJ. An aminopeptidase, ARTS-1, is required for interleukin-6 receptor shedding. *J. Biol. Chem.* 2003; 278:28677–28685. [PubMed: 12748171]
25. Cui X, et al. Identification of ARTS-1 as a novel TNFR1-binding protein that promotes TNFR1 ectodomain shedding. *J. Clin. Invest.* 2002; 110:515–526. [PubMed: 12189246]
26. Evnouchidou I, Berardi MJ, Stratikos E. A continuous fluorogenic assay for the measurement of the activity of endoplasmic reticulum aminopeptidase 1: competition kinetics as a tool for enzyme specificity investigation. *Anal. Biochem.* 2009; 395:33–40. [PubMed: 19638272]
27. Libioulle C, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of *PTGER4*. *PLoS Genet.* 2007; 3:e58. [PubMed: 17447842]
28. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat. Rev. Immunol.* 2008; 8:458–466. [PubMed: 18500230]
29. Zhang XJ, et al. Psoriasis genome-wide association study identifies susceptibility variants within *LCE* gene cluster at 1q21. *Nat. Genet.* 2009; 41:205–210. [PubMed: 19169255]
30. Cargill M, et al. A large-scale genetic association study confirms *IL12B* and leads to the identification of *IL23R* as psoriasis-risk genes. *Am. J. Hum. Genet.* 2007; 80:273–290. [PubMed: 17236132]
31. Duerr RH, et al. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science.* 2006; 314:1461–1463. [PubMed: 17068223]
32. 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. [PubMed: 19479869]
33. Gagliardi MC, et al. Endogenous PGE2 promotes the induction of human Th17 responses by fungal  $\beta$ -glucan. *J. Leukoc. Biol.* 2010; 88:947–954. [PubMed: 20807707]
34. Ruutu M, et al. Fungal  $\beta$ -glucan triggers spondyloarthropathy and Crohn's disease in SKG mice. *Arthritis Rheum.* 2010; S1:1446.
35. Hirota K, et al. T cell self-reactivity forms a cytokine milieu for spontaneous development of IL-17+ Th cells that cause autoimmune arthritis. *J. Exp. Med.* 2007; 204:41–47. [PubMed: 17227914]
36. Armaka M, et al. Mesenchymal cell targeting by TNF as a common pathogenic principle in chronic inflammatory joint and intestinal diseases. *J. Exp. Med.* 2008; 205:331–337. [PubMed: 18250193]

37. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am. J. Hum. Genet.* 1987; 40:1–14. [PubMed: 3468804]
38. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* 1990; 46:222–228. [PubMed: 2301392]
39. Strange A, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAPI1*. *Nat. Genet.* 2010; 42:985–990. [PubMed: 20953190]
40. Jaakkola E, et al. Finnish HLA studies confirm the increased risk conferred by HLA-B27 homozygosity in ankylosing spondylitis. *Ann. Rheum. Dis.* 2006; 65:775–780. [PubMed: 16249228]
41. Rudwaleit M, et al. Effectiveness, safety, and predictors of good clinical response in 1250 patients treated with adalimumab for active ankylosing spondylitis. *J. Rheumatol.* 2009; 36:801–808. [PubMed: 19273449]
42. Pointon JJ, et al. The chromosome 16q region associated with ankylosing spondylitis includes the candidate gene tumour necrosis factor receptor type 1-associated death domain (*TRADD*). *Ann. Rheum. Dis.* 2010; 69:1243–1246. [PubMed: 19854717]
43. Conrad DF, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010; 464:704–712. [PubMed: 19812545]
44. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics.* 2007; 23:2741–2746. [PubMed: 17846035]
45. Spencer CC, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of *SNCA* and multiple associated haplotypes at 17q21. *Hum. Mol. Genet.* 2011; 20:345–353. [PubMed: 21044948]
46. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; 5:e1000529. [PubMed: 19543373]
47. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* 2009; 10:387–406. [PubMed: 19715440]
48. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81:559–575. [PubMed: 17701901]
49. R Core Development Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna: 2009.
50. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008; 24:1547–1548. [PubMed: 18467348]
51. Simon R, et al. Analysis of gene expression data using BRB-array tools. *Cancer Inform.* 2007; 3:11–17. [PubMed: 19455231]



**Figure 1.**

Association findings for the *ERAP1* SNP rs30187 stratified by the HLA-B27 tag SNP rs4349859. rs4349859 allele A tags HLA-B27. Error bars, 95% confidence intervals. Note that the *ERAP1* risk allele T only increases risk in individuals carrying at least one copy of the HLA-B27 risk allele tag. The odds ratios for the genotype combinations were calculated using logistic regression in the R software package including covariates for ancestry (where appropriate). Genotypes with the low risk CC/GG genotype were set as the baseline, and the other genotype combinations were coded according to a series of dichotomous indicator variables. Odds ratios were derived by exponentiating the relevant coefficient from the logistic regression.



**Figure 2.**

The mean rate of trimming of N-terminal tryptophan (mol, substrate/mol, enzyme/sec) from 10-mer peptide WRVYEKCALK by wild-type ERAP1 and variants associated with ankylosing spondylitis (rs30187 (p.Lys528Arg), rs17482078 (p.Arg725Gln) and rs10050860 (p.Asp575Asn)). WT, wild type. Circles represent results for wild-type genotype samples, squares represent results for rs30187, upright triangles represent results for rs17482078 and inverted triangles refer to results for rs10050860.

Table 1

Association evidence at loci previously identified as influencing ankylosing spondylitis risk

Chr.	SNP ID	Position <sup>a</sup>	Putative gene	Risk allele	WTCCC2 discovery sample			TASC discovery sample			Replication sample			Combined
					P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	
1p31	rs11209026	67,478,546	<i>IL23R</i>	G	$1.9 \times 10^{-6}$	1.48	1.26–1.73	$2.8 \times 10^{-7}$	1.90	1.49–2.43	$4.2 \times 10^{-7}$	1.61	1.34–1.93	$2.3 \times 10^{-17}$
1q32	rs2297909	199,226,930	<i>KIF21B</i>	G	0.0017	1.14	1.05–1.24	0.00034	1.22	1.09–1.35	$1.8 \times 10^{-7}$	1.25	1.15–1.36	$5.2 \times 10^{-12}$
2p15	rs10865331	62,404,976	–	A	$2.9 \times 10^{-12}$	1.32	1.22–1.43	$1.7 \times 10^{-9}$	1.34	1.22–1.47	$2.5 \times 10^{-15}$	1.36	1.26–1.47	$6.5 \times 10^{-34}$
5q15	rs30187	96,150,086	<i>ERAP1</i>	T	$4.3 \times 10^{-13}$	1.35	1.24–1.46	$7.1 \times 10^{-8}$	1.31	1.19–1.44	$1.1 \times 10^{-9}$	1.28	1.18–1.39	$1.8 \times 10^{-27}$
6p21	rs4349859	31,473,766	<i>HLA-B</i>	A	$<10^{-200}$	48.9	43.3–55.2	$<10^{-200}$	90.4	72.3–113	$<10^{-200}$	40.8	35.0–47.6	$<10^{-200}$
21q22	rs378108	39,391,390	–	G	$1.5 \times 10^{-7}$	1.23	1.14–1.33	$2.9 \times 10^{-6}$	1.25	1.14–1.38	0.059	1.08	1.00–1.16	$2.1 \times 10^{-11}$

See text for references. HLA-B\*27 status is sometimes used as a diagnostic tool (especially in the TASC data), and this ascertainment effect means that estimates of its odds ratio (OR) may be biased upwards. Chr., chromosome.

<sup>a</sup>NCBI human genome build 36 coordinates.



Table 2

Newly confirmed associations (upper tier) and those with strong and replicated evidence for association (lower tier)

Chr.	SNP ID	Position <sup>a</sup>	Putative gene	Risk allele	WTCCC2 discovery sample				TASC discovery sample				Replication sample				Combined	
					P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	
1p36	rs11249215	25,169,771	<i>RUNX3</i>	A	$2.1 \times 10^{-5}$	1.19	1.10–1.28	0.00091	1.19	1.07–1.32	0.00029	1.15	1.07–1.24	$9.2 \times 10^{-11}$				
5q33	rs6556416	158,751,323	<i>IL12B</i>	C	$1.0 \times 10^{-4}$	1.18	1.08–1.28	0.074	1.11	0.99–1.23	0.00018	1.17	1.08–1.27	$1.9 \times 10^{-8}$				
12p13	rs11616188	6,373,003	<i>LTBR-TNFRSF1A</i>	A	$1.2 \times 10^{-5}$	1.21	1.11–1.31	$6.2 \times 10^{-8}$	1.38	1.23–1.56	0.0036	1.12	1.04–1.21	$4.1 \times 10^{-12}$				
4q21	rs4389526	81,165,499	<i>ANTXR2</i>	A	0.00054	1.15	1.06–1.25	0.00027	1.21	1.09–1.34	0.019	1.10	1.02–1.19	$9.4 \times 10^{-8}$				
5p13	rs10440635	40,526,547	<i>PTGER4</i>	A	0.0027	1.13	1.04–1.22	0.00020	1.20	1.09–1.32	0.017	1.10	1.02–1.19	$2.6 \times 10^{-7}$				
9q34	rs10781500	138,389,159	<i>CARD9</i>	T	$2.0 \times 10^{-5}$	1.18	1.09–1.28	0.043	1.10	1.00–1.22	0.041	1.08	1.00–1.17	$1.1 \times 10^{-6}$				
17q21	rs8070463	43,123,835	<i>TBKBP1</i>	C	0.0025	1.13	1.04–1.22	$1.6 \times 10^{-5}$	1.24	1.12–1.37	0.015	1.10	1.02–1.18	$5.3 \times 10^{-8}$				

Top tier SNPs have a combined  $P < 5 \times 10^{-8}$  and lower tier SNPs have combined  $P = 5 \times 10^{-6}$ . Chr., chromosome.

<sup>a</sup>NCBI human genome build 36 coordinates.

Table 3

Evidence for secondary signals of association at *IL23R*, *ERAP1* and *HLA-B27*

Chr.	SNP ID	Position <sup>a</sup>	Conditional on SNP	Effect allele	WTCCC2 discovery sample			TASC discovery sample			Replication sample			Combined		
					P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI
1p31	rs11209032	67,512,680	rs11209026	A	$7.9 \times 10^{-6}$	1.20	1.11–1.30	0.0018	1.17	1.06–1.29	0.0018	1.14	1.05–1.23	$5.1 \times 10^{-10}$		
5q15	rs10050860	96,147,966	rs30187	C	0.0025	1.17	1.06–1.30	0.026	1.16	1.02–1.33	0.00023	1.22	1.10–1.35	$1.9 \times 10^{-7}$		
6p21	rs13202464	31,452,562	rs4349859 <sup>b</sup>	A	$3.7 \times 10^{-7}$	0.32	0.21–0.47	$2.1 \times 10^{-15}$	0.15	0.09–0.23	$2.0 \times 10^{-9}$	0.34	0.25–0.47	$1.1 \times 10^{-29}$		

<sup>a</sup>NCBI human genome build 36 coordinates.<sup>b</sup>We assumed a dominant model for the risk allele at this SNP.