Human Molecular Genetics, 2009, Vol. 18, No. 21 **4204–4212**
doi:10.1093/hmg/ddp371
Advance Access published on August 18, 2009

Investigating the genetic association between *ERAP1* and ankylosing spondylitis

David Harvey[1], Jennifer J. Pointon[1], David M. Evans[2], Tugce Karaderi[1], Claire Farrar[1], Louise H. Appleton[1], Roger D. Sturrock[3], Millicent A. Stone[4], Udo Oppermann[5], Matthew A. Brown[1,6] and B. Paul Wordsworth[1,*]

[1]Botnar Research Centre, Institute of Musculoskeletal Science, Oxford OX37LD, UK, [2]MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol BS8 2BN, UK, [3]Centre for Rheumatic Diseases, Glasgow Royal Infirmary, Glasgow G31 2ER, UK, [4]Royal National Hospital for Rheumatic Diseases, Bath BA1 1RL, UK, [5]Structural Genomics Consortium, Old Road Campus Research Building, Oxford OX3 7DQ, UK and [6]Diamantina Institute of Cancer, Immunology and Metabolic Medicine, University of Queensland, Brisbane, QLD 4102, Australia

Received March 10, 2009; Revised and Accepted August 3, 2009

A strong association between *ERAP1* and ankylosing spondylitis (AS) was recently identified by the Wellcome Trust Case Control Consortium and the Australo-Anglo-American Spondylitis Consortium (WTCCC-TASC) study. *ERAP1* is highly polymorphic with strong linkage disequilibrium evident across the gene. We therefore conducted a series of experiments to try to identify the primary genetic association(s) with *ERAP1*. We replicated the original associations in an independent set of 730 patients and 1021 controls, resequenced *ERAP1* to define the full extent of coding polymorphisms and tested all variants in additional association studies. The genetic association with *ERAP1* was independently confirmed; the strongest association was with rs30187 in the replication set ($P = 3.4 \times 10^{-3}$). When the data were combined with the original WTCCC-TASC study the strongest association was with rs27044 ($P = 1.1 \times 10^{-9}$). We identified 33 sequence polymorphisms in *ERAP1*, including three novel and eight known non-synonymous polymorphisms. We report several new associations between AS and polymorphisms distributed across *ERAP1* from the extended case–control study, the most significant of which was with rs27434 ($P = 4.7 \times 10^{-7}$). Regression analysis failed to identify a primary association clearly; we therefore used data from HapMap to impute genotypes for an additional 205 non-coding SNPs located within and adjacent to *ERAP1*. A number of highly significant associations ($P < 5 \times 10^{-9}$) were identified in regulatory sequences which are good candidates for causing susceptibility to AS, possibly by regulating *ERAP1* expression.

INTRODUCTION

Ankylosing spondylitis (AS) is a highly heritable form of spondyloarthropathy with an oligogenic component to susceptibility. The best known contribution to AS comes from the HLA-B27 immune response gene, accounting for ~37% of the total genetic risk (1). HLA-B27 is virtually a prerequisite for the development of AS but on its own it is insufficient to cause the disease (1). In 2007, the Wellcome Trust Case Control Consortium and the Australo-Anglo-American Spondylitis Consortium (WTCCC-TASC) performed a gene-targeted association study of ~14 500 nsSNPs in 1000 AS cases and 1500 controls. They confirmed the known strong association with the major histocompatibilty complex (MHC) and also provided preliminary evidence for several non-MHC associations. Five nsSNPs in *ERAP1* were genotyped, all producing significant associations with AS with a peak *P*-value of 1×10^{-6} (2). The association with endo-

*To whom correspondence should be addressed at: Botnar Research Centre, University of Oxford Institute of Musculoskeletal Science, Windmill Road, Headington, Oxford OX3 7LD, UK. Tel: +44 1865737545; Fax: +44 1865737640; Email: paul.wordsworth@ndm.ox.ac.uk

plasmic reticulum associated aminopeptidase 1 (*ERAP1*), also known as *ARTS1*, was subsequently replicated in an independent North American cohort and recently further confirmation has come from case–control and familial association studies (2–4).

ERAP1 encodes a multi-functional zinc-metallopeptidase belonging to the M1 family of aminopeptidases with several proposed biological functions that make it a strong candidate in AS. ERAP1 has been shown to bind directly to the extracellular domain of TNFR1 *in vitro* and promote its IL-1β-mediated ectodomain cleavage to generate 27–34 kDa soluble TNFR1 (5). Cleavage may require ERAP1 to be in a calcium-dependent complex with nucleobindin 2 (NUCB2) and RNA-binding motif protein, X-linked (RBMX) (6–9). This ERAP1–NUCB2–RBMX complex is also required for the constitutive release of full length 55 kDa TNFR1 within exosome-like vesicles to the extracellular compartment (6,7). Such ERAP1 assisted generation of extracellular TNFR1 could be crucial for regulating the bioactivity of TNF which plays a central role in regulation of inflammation. A direct correlation has also been demonstrated between increased membrane-associated ERAP1 protein and increased soluble IL-6Rα and IL-1RII, consistent with their ectodomain cleavage (9,10).

In the endoplasmic reticulum (ER), ERAP1 has the potential to trim peptide antigens to optimal length for binding to MHC class I molecules (11). Complex proteins are initially degraded in the cytosol by the multi-unit proteasome complex, typically generating peptide fragments up to 25 amino acids in length (12). Such peptide antigens and their N-terminal extended precursors are subsequently transported into the ER by the transporter associated with antigen processing that preferentially transports peptides 8–16 residues in length (13–16). Nascent MHC class I molecules typically bind short peptide fragments 8–9 residues long and transport them to the cell surface for presentation to T cells. ERAP1 is expressed in the lumen of the ER where peptide loading to MHC class I molecules takes place. Here it preferentially trims substrates 10–16 residues long whilst sparing peptides 8–9 residues in length, the optimal length for binding MHC class I molecules (17,18).

Determining the precise role of ERAP1 in susceptibility to AS is now a major research goal. There have now been several studies indicating association between AS and ERAP1 but no fine-mapping studies have previously been reported. Here we report the first UK replication in a large independent cohort of cases and controls. Subsequently, we have systematically identified all common *ERAP1* coding variants present in AS patients and genotyped them in a large extended case–control study. We report a number of new associations and have attempted to correlate the strongest associations with potentially important functions of the ERAP1 protein and its expression.

RESULTS

Independent ERAP1 replication study

We genotyped four SNPs from the initial WTCCC-TASC study in an independent cohort of 730 AS cases and 1021 controls. All *ERAP1* SNPs were significantly associated with AS (Table 1). This data was then combined with that from the WTCCC-TASC study (1000 AS cases and 4379 controls) and a meta-analysis performed. All SNPs achieved a greater level of significance with the most significant association seen with rs30187 in the replication study ($P = 3.4 \times 10^{-3}$) and with rs27044 in the combined data set ($P = 1.1 \times 10^{-9}$).

ERAP1 sequencing

In the 48 individuals with AS who were sequenced we observed 33 sequence variants shown in Table 2, 28 of which had been previously reported; 14 were known coding SNPs, three were SNPs located close to exon–intron boundaries (within 25 nucleotides of an exon) and two SNPs were identified in the 5′-UTR predicted to encode the *ERAP1* promoter (19). The remaining nine known sequence variants were intronic (within 40–100 nucleotides of an exon) or downstream of the STOP codon in exon 20. We also found five new sequence variants in *ERAP1*, three of which are non-synonymous variants in highly conserved amino acids present in single individuals. The final two novel variants observed were intronic (variants 2 and 5). Five rare coding HapMap SNPs [minor allele frequencies (MAF) < 0.04] were not observed in this study.

Extended *ERAP1* case control study

We genotyped 23 *ERAP1* polymorphisms [excluding the sequence variants exhibiting strong linkage disequilibrium (LD)] in 1604 AS cases and 1021 healthy controls. As expected (due to the overlap of AS cases and controls with previous studies), we observed a strong association between AS and the five SNPs (rs30187, rs27044, rs2887987, rs17482078 and rs10050860) reported in the initial WTCCC-TASC study (Table 3). We also observed highly significant ($P < 9 \times 10^{-5}$) associations with four additional *ERAP1* SNPs (rs26653, rs27434, rs1065407 and rs13167972), the most strongly associated being rs27434 (4.7×10^{-7}) and rs1065407 (7.5×10^{-7}). We also report weak associations with two other SNPs rs28366066 ($P = 0.009$) and rs469876 ($P = 5.9 \times 10^{-4}$), all of which were not part of the original WTCCC-TASC study. Two of the new variants identified by sequencing (see novel variants 3 and 4 in Tables 2 and 3) were only observed in patients with AS: four AS patients were heterozygous for novel variant 3, resulting in a valine/isoleucine substitution; two other AS patients were heterozygous for novel variant 4, resulting in a substitution of arginine for cysteine.

Conditional logistic regression was performed on the six SNPs in *ERAP1* (rs27434, rs1065407, rs30187, rs26653, rs27044 and rs17482078) that were most strongly associated with AS in the extended study. This was to test whether any individual SNP was independently associated with AS after controlling for associations arising from the other five SNPs on the haplotype. When all the other five markers were controlled for individually in this way, no single marker showed obvious residual association ($P > 0.1$) (Table 4A). We also tested whether controlling for the association arising from each SNP individually left any residual association with the remaining haplotype. In this case, for each marker tested individually, there was strong residual haplotypic association

Table 1. Cochrane–Armitage test of trend for replication cohort and meta-analysis for combined cohort (WTCCC-TASC UK AS cases plus replication AS cases compared with 1958 British Birth Cohort, non-AS WTCCC disease groups and replication controls)

SNP	Minor allele	Replication cohort				Combined cohort			
		Case MAF ($n = 730$)	Control MAF ($n = 1021$)	OR	*P*-value	Case MAF ($n = 1730$)	Control MAF ($n = 5400$)	OR	*P*-value
rs27044	G	0.31	0.27	1.2	3.9×10^{-3}	0.33	0.27	1.3	1.1×10^{-9}
rs10050860	T	0.19	0.23	0.8	6.7×10^{-3}	0.18	0.23	0.8	9×10^{-8}
rs30187	T	0.39	0.34	1.2	3.4×10^{-3}	0.40	0.34	1.3	5×10^{-9}
rs2287987	C	0.18	0.23	0.8	3.6×10^{-3}	0.18	0.23	0.8	5.4×10^{-8}

Table 2. *ERAP1* sequence variants identified by sequencing, including their location and amino acid change in 48 AS cases

SNP	Minor allele	Position	Location/ exon	Amino acid change
rs28366066	A	96169981	5′-UTR	Intronic
rs151949	G	96169601	5′-UTR	Intronic
rs26653	C	96165006	2	Pro127Arg
rs27528	T	96162281	IVS3+40	Intronic
rs26618	C	96156592	5	Ile276Met
Novel variant 1	T	96156570	5	Asp284Asn
rs27895	T	96155299	6	Gly346Asp
rs2287987	C	96155291	6	Met349Val
rs27434	A	96155268	6	Ala356
rs27640	T	96153661	IVS8−10	Intronic
rs3213809	A	96153589	8	His417
rs27529	A	96152064	9	Ser453
rs11743410	G	96151015	IVS11−65	Intronic
rs30186	G	96150203	IVS11−59	Intronic
rs30187	T	96150086	11	Lys528Arg
rs30379	T	96148016	IVS12−6	Intronic
rs10050860	T	96147966	12	Asp575Asn
Novel variant 2	C	96147891	IVS12+39	Intronic
rs469758	C	96147471	IVS13−40	Intronic
rs469783	C	96147280	13	Ala637
Novel variant 3	T	96147252	13	Val647Ile
rs469876	G	96147162	IVS13+85	Intronic
rs17482078	T	96144622	15	Arg725Gln
rs27044	G	96144608	15	Gln730Glu
Novel variant 4	G	96144590	15	Cys736Arg
ENSSNP12878388	T	96143311	16	Leu763
Novel variant 5	C	96143114	IVS16+40	Intronic
rs17481856	A	96142564	17	Leu848
rs1065407	G	96137839	IVS19+25	Intronic
rs3822682	G	96123970	20	Non-coding
rs27582	A	96123651	20	Non-coding
rs27980	G	96122584	20	Non-coding
rs13167972	G	96142564	20	Non-coding

($P < 0.001$) (Table 4B). Thus although these data show that each of these SNPs are associated with AS they cannot distinguish any particular individual SNP as being primarily responsible for the association at the locus, probably because of the tight LD across the locus. The data are consistent either with there being a single predominant associated variant or with existence of multiple associated alleles. Our data have insufficient power to distinguish between these two possibilities.

Therefore, we imputed genotypes for an additional 205 SNPs within and surrounding *ERAP1* using our genotype data and data from HapMap (Supplementary Material, 1) to assess the level of association across this region. We observed many additional SNPs strongly associated with AS from the imputation, including 16 SNPs which obtained a higher level of significance than any SNP from our case–control study (Fig. 1). The most significant associations were with a rare intron 19 SNP, rs27689 (96 132 957 bp, MAF = 0.07, imputation quality score 0.88, $P = 3.1 \times 10^{-9}$) and a group of intronic SNPs in very strong LD, the most significant of which was the intron 16 SNP rs27043 (96 143 055 bp, imputation quality score 0.99, $P = 3.2 \times 10^{-9}$).

Comparison with *ERAP1 cis*-SNP expression

P-values for imputed/genotyped SNPs associations with AS in our study were compared with *P*-values for association with *ERAP1* expression from the previous study (20). Only a subset of 32 SNPs were used in both studies but these data showed a correlation (using probe 209788_s_at) between the strength of SNP association with AS and strength of association with *ERAP1* expression ($r = 0.75$, $P < 0.0001$) (Fig. 2). The SNPs showing greatest association with *ERAP1* expression from the previous study were rs28096 (1.6×10^{-32}) and rs27434 (2.7×10^{-32}) which are also strongly associated with AS in our extended case–control/imputation study (7.8×10^{-8} and 4.7×10^{-7}, respectively).

Molecular modeling of ERAP1 protein structure

Molecular modeling of ERAP1 suggests a three-domain protein structure surrounding a central Zn atom. Only rs2287987 (Met349Val) was located close to the HEXXH-(X_{18})-E zinc-binding motif whereas the remaining variants were dispersed widely across the predicted structure (Fig. 3).

DISCUSSION

We have robustly confirmed the genetic association between AS and *ERAP1*, first in an independent replication study in UK subjects and subsequently with additional markers in a further case–control study. In addition to 14 known coding polymorphisms, we identified three novel coding variants in *ERAP1*, two of which were observed only in AS patients. Finally, we found six new associations with SNPs in *ERAP1* with AS, and many additional markers by imputation which are possible candidates for the causative variant(s) in this gene.

The strongest associations identified were with rs27434 (4.7×10^{-7}) in the extended case–control study and with rs27689 (3.1×10^{-9}) and rs27043 (3.2×10^{-9}) by imputa-

Table 3. Case and control MAF, odds ratios (OR), upper and lower 95% confidence limits (95% CI), *P*-values and the power to detect an association to an OR of 1.2 are shown for the SNPs genotyped in the *ERAP1* case–control study

SNP	Case MAF ($n = 1604$)	Control MAF ($n = 1021$)	OR	95% CI	*P*-value	Power to detect
rs28366066	0.06	0.08	0.74	1.07–1.70	0.009	41%
rs151949	0.37	0.39	0.92	0.96–1.22	0.17	89%
rs26653	0.33	0.27	1.3	0.68–0.87	8.4×10^{-5}	86%
rs27582	0.26	0.25	1.02	0.79–1.28	0.59	82%
rs26618	0.23	0.24	0.99	0.79–1.27	0.97	81%
NV1	0.00	0.00	N/A	N/A	N/A	N/A
rs27895	0.06	0.07	1.07	0.85–1.34	0.48	36%
rs2287987	0.18	0.23	0.76	1.15–1.52	1.4×10^{-4}	83%
rs27434	0.26	0.20	1.44	0.6–0.8	4.7×10^{-7}	82%
rs3213809	0.15	0.15	0.96	0.8–1.15	0.76	67%
rs30187	0.40	0.34	1.32	0.67–0.85	4.9×10^{-6}	89%
rs10050860	0.18	0.23	0.75	1.16–1.54	1.2×10^{-4}	83%
NV2	0.005	0.004	1.23	0.17–3.8	0.6	7%
rs469783	0.47	0.42	1.22	1.08–1.37	5.9×10^{-4}	90%
NV3	0.001	0	N/A	N/A	0.1	N/A
rs469876	0.25	0.25	1.01	0.89–1.16	0.72	82%
rs27044	0.34	0.27	1.36	0.65–0.83	1.6×10^{-6}	86%
rs17482078	0.18	0.23	0.73	1.19–1.59	4×10^{-5}	83%
NV4	0.003	0	N/A	N/A	0.25	N/A
ENSSNP12878388	0.06	0.08	0.79	1.01–1.6	0.05	41%
NV5	0.01	0.01	1.19	0.24–1.76	0.6	10%
rs17481856	0.14	0.14	0.98	0.84–1.24	0.81	65%
rs1065407	0.29	0.36	0.74	1.20–1.53	7.5×10^{-7}	87%
rs27980	0.32	0.32	1.02	0.9–1.17	0.7	87%
rs13167972	0.33	0.39	0.74	1.17–1.51	9.1×10^{-6}	89%

Table 4. Conditional logistic regression analysis of most associated ERAP1 loci

(A) Lack of association with individual named *ERAP1* markers after controlling for association with the other five markers. This could not be tested at rs30187 or rs27434 because they are in complete linkage disequilibrium in this dataset

Marker	Residual association
rs1065407	0.24
rs27044	0.28
rs30187	–
rs27434	–
rs31087–rs27434	0.36
rs2287987	0.6
rs26653	0.24

(B) Residual *ERAP1* haplotypic association after controlling for the association of the named marker

Marker	Residual association
rs1065407	3.4×10^{-5}
rs27044	0.00015
rs30187	0.00016
rs27434	0.00098
rs2287987	3.7×10^{-6}
rs26653	3.4×10^{-5}

tion. As a result of the extensive LD within and surrounding *ERAP1*, associations were observed throughout the ~145 kb investigated. Consequently locating the primary association will be extremely difficult using a solely genetic approach in white European cohorts alone (Fig. 1). Conditional logistic regression analysis performed on the six most significant genotyped markers from the extended case–control study (which focused on coding SNPs and potential splice-site variants) did not allow us to identify an obvious primary association. However, a contribution to a functional effect by any of these SNPs on ERAP1 cannot be completely excluded from this analysis alone. The polymorphism rs27434 contributes to the conserved HEXXH-(X_{18})-E zinc-binding motif of ERAP1 and the modeled protein structure of ERAP1 suggests rs2287987 is also in close proximity to the catalytic domain (Fig. 3). However, rs27434 seems extremely unlikely to have any effect on its activity because it is a synonymous polymorphism. Goto and colleagues reported that rs30187 caused a significant reduction in aminopeptidase activity towards a synthetic peptide substrate, resulting in 2–3-fold decrease in substrate affinity. Using these data and molecular modeling of the putative ERAP1 substrate pocket they hypothesized that this residue contributes to substrate-binding affinity with the enzyme (21). Finally, rs1065407 is located 17 nucleotides 3′ of the exon 19 splice site that results in a premature stop codon in isoform 1 of ERAP1. A number of the most strongly associated SNPs from the imputed data also reside within introns 18 and 19. The importance of abnormalities of splicing in causing disease has been highlighted recently (22,23).

Recent studies suggest that *ERAP1* expression traits may be under *cis*-regulatory control of expression (20,24,25). Such heritable allelic expression is common throughout the human genome (26). We identified a correlation between the strength of *ERAP1* SNPs association with AS and strength of the same SNPs association with ERAP1 transcript abundance from a previous expression study (20) This suggests *cis*-SNP control of *ERAP1* expression may be of importance in AS, but expression studies that relate these SNPs to individual

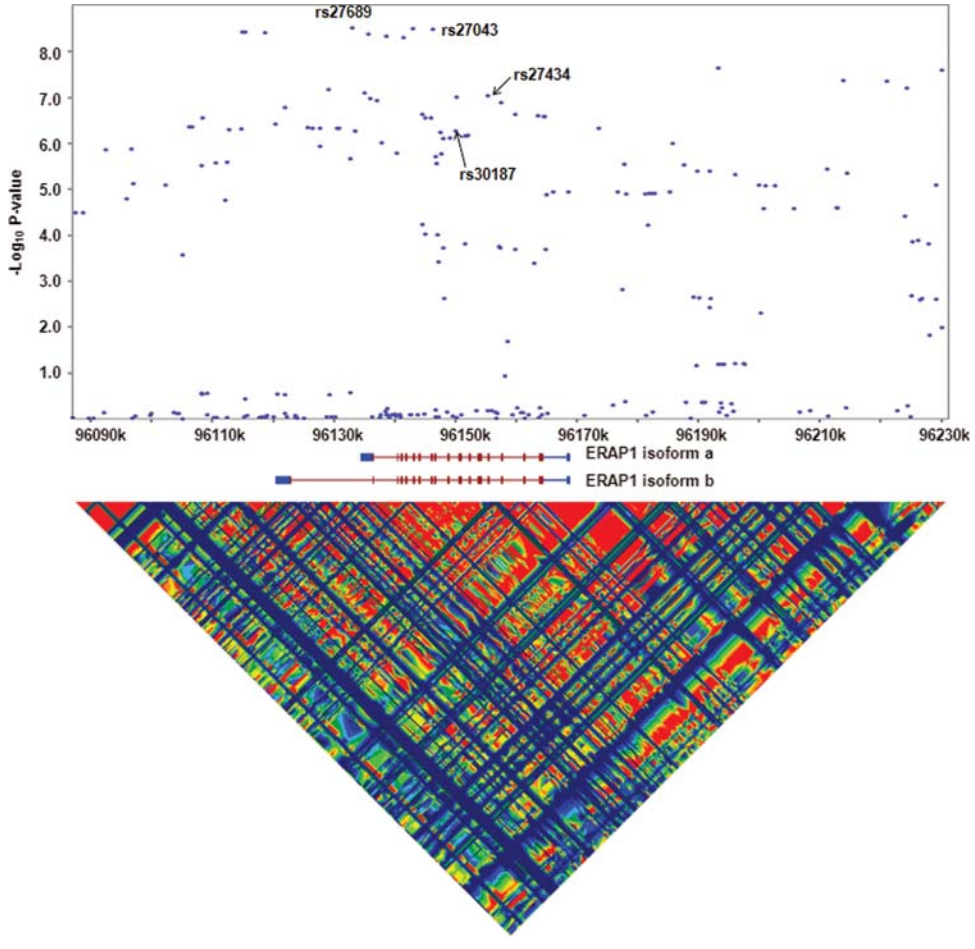

Figure 1. SNP association with AS against genomic location. Log_{10} *P*-value for imputed and genotyped SNPs plotted against physical distance. Aligned to the plot are the genomic location of two ERAP1 isoforms and LD that exists in this region, red regions indicating high LD and blue indicating low LD.

transcript levels are required to fully elucidate the genetic factors that control of *ERAP1* expression. The potential implication of several *ERAP1* nsSNPs in AS also underlines the importance of solving its corresponding three-dimensional protein structure which could be critical to understanding how these polymorphisms affect its structure, function and ability to interact with binding partners.

Two novel potentially disease-associated mutations were identified exclusively in AS patients (V647I and C736R). It is most likely that these variants were not observed in the control population purely by chance but there is evidence that similar private mutations in *CARD15/NOD2* may be associated with Crohn's disease (27). In the absence of functional studies illustrating the genotype–phenotype relationship, validating such private mutations is extremely difficult because of the difficulty of conducting sufficiently powered association studies.

There are two major putative functions of ERAP1 that could explain its role in AS but our data do not help to distinguish between them. These functions involve the cleavage of cell-surface cytokine receptors (e.g. TNFR1) and peptide binding to MHC class I molecules in the ER. TNFR1 shedding by ERAP1 may decrease TNF bioactivity not only by reducing the number of available cell-surface receptors to induce TNFR1-mediated signaling but also by increasing the amount of soluble TNFR1 that competes with cell-surface TNFR1 for extracellular TNF (28,29). Loss of function ERAP1 variants could clearly result in a pro-inflammatory state via these mechanisms. However, ERAP1 is not the 'sheddase' responsible for actually cleaving TNFR1 *in vitro* and there is a need for additional studies to provide supporting evidence for this function of ERAP1. In contrast, the peptide trimming function of ERAP1 is well described. ERAP1 has been isolated from the ER, where peptide binding to the nascent MHC class I molecules occurs, and iRNA studies suggest it is involved in the formation of about one-third of MHC-peptide complexes (9,30). Functional studies of variant forms of ERAP1 could help assess the role of this protein in peptide loading in patients with AS and the general population.

To conclude, we have confirmed the genetic association between *ERAP1* and AS and have identified a range of polymorphisms distributed throughout, and surrounding the gene that are associated with AS. The most strongly associated SNPs identified in this study by imputation reside within *ERAP1* regulatory sequences. Only with subsequent studies to determine the impact of these polymorphisms on expression and their structural and functional consequences is it likely that the nature of the association with AS will finally be determined.

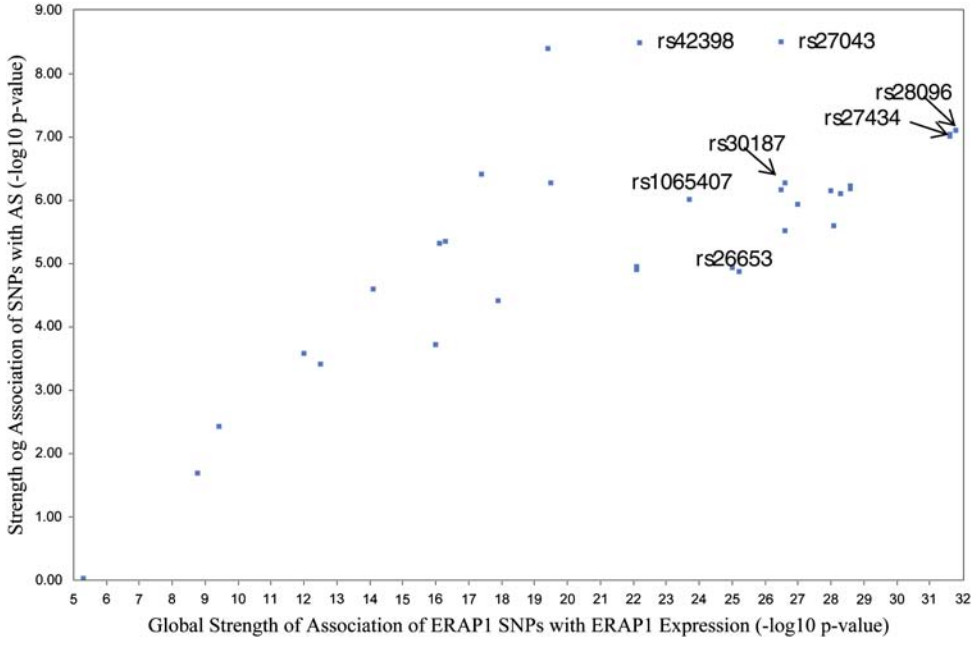

Figure 2. Comparison of the strength of association of various *ERAP1* SNPs with AS and their influence on ERAP1 expression. *P*-values for association of SNPs with AS generated by genotyping or imputation in our extended cases–control study were compared with *P*-values for SNP association with ERAP1 transcript abundance from a previous study (using data from probe 209788_s_at) (18). Only a subset of 32 SNPs are plotted which were used in both studies.

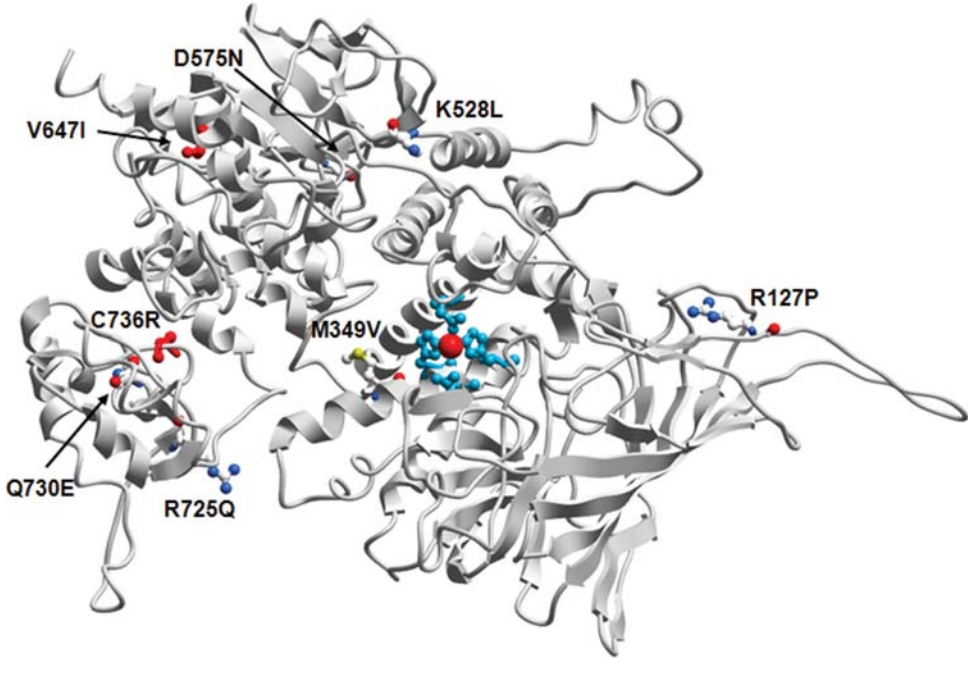

Figure 3. Ribbon model of putative ERAP1 structure. Highlighted are the active site around the central Zn atom (red sphere) in cyan and the positions of residues affected by non-synonymous polymorphisms associated with AS in this study, including rare novel variants V647I and C736R.

MATERIALS AND METHODS

As cases and controls

All patients with AS were recruited with informed consent (MREC project number 98/5/23) from patients attending the Nuffield Orthopaedic Centre (Oxford, UK), the Royal National Hospital for Rheumatic Diseases (Bath, UK), the Centre for Rheumatic Diseases (Glasgow, UK), members of the UK National Ankylosing Spondylitis Society (NASS) or by referral from other British rheumatologists. All patients were

Caucasians, of UK origin and fulfilled the modified New York criteria for AS (31). DNA was prepared from peripheral blood leucocytes by standard methods. For comparison in both the replication and fine-mapping extended study we used 1021 ethnically matched healthy subjects, who were either blood donors or spouses of patients with osteoarthritis.

Independent *ERAP1* replication study

We genotyped four SNPs in *ERAP1* in 730 unrelated AS cases and 1021 healthy controls by iPLEX technology, which is based on a single base pair primer extension followed by mass spectrometry (MassARRAY®, Sequenom, San Diego, USA). Case and control genotype frequencies were compared and statstical significance was tested using the Cochrane–Armitage test of trend. Finally, using the StatsDirect statistical package (http ://www.statsdirect.com. England: StatsDirect Ltd. 2008) we tested for heterogeneity between our replication cohort and the WTCCC-TASC data (1000 AS cases compared with 1500 controls from the 1958 British Birth Cohort and 2879 disease control individuals with multiple sclerosis, autoimmune thyroid disease or breast cancer which are not associated with *ERAP1*) using the Cochran Q statistic. Subsequently, based on an insignificant Cochran Q *P*-value, a meta-analysis was performed using the Mantel–Haenszel test for fixed effects OR.

ERAP1 sequencing

The 20 exons, intron–exon boundaries and 5′-UTR of *ERAP1* were sequenced in 48 new AS cases. The consensus sequence was obtained from Ensembl Genome Browser (http://www.ensembl.org/). Primers were designed for PCR products of 400–550 bp with overlapping PCR products for target regions >550 bp. Sequencing was carried out using BigDye v3.1 (Applied Biosystems, Warrington, UK) followed by purification using Pellet Paint® Co-Precipitant (Novagen, Northumberland, UK). Finally, samples were added to HiDi formamide and analyzed using an AB3100 Genetic Analyzer (Applied Biosystems, Warrington, UK). All sequence variants were identified by manual inspection of the electropherograms using the software program SeqScape v2.1. *ERAP1* sequence data were used to calculate statistical measures of LD between SNPs using the LDMAX programme (32).

Extended *ERAP1* case–control study

Twenty-three SNPs (six tagging SNPs, with $r^2 > 0.8$ identified from the *ERAP1* sequencing data, the remaining nine sequence variants, two SNPs in the 5′-UTR and the six novel sequence variants) were selected for genotyping in 1604 AS cases (a combination of AS cases included in the original WTCCC-TASC study and our independent replication and sequencing studies) and 1021 controls (used in the replication study). SNPs were genotyped by iPLEX technology and statistical significance was tested using the Cochrane–Armitage test of trend using the PLINK program (33). Assays were initially run using CEPH HapMap individuals to confirm they were functioning correctly and their genotype clusters used as a reference for the test subjects. Markers that failed Hardy–Weinberg equilibrium in controls or were missing >10% of genotypes were excluded from analyses. We also manually checked genotype calls for each SNP by inspecting the cluster plots to ensure three distinct clusters were apparent and any outliers excluded. The power to detect an association to an OR of 1.2 was calculated using a log additive model, assuming prevalence of 0.4%, observed case and control allele frequencies, and an alpha level of 0.05. Conditional logistic regression analyses involving the six most significantly associated SNPs were performed using PLINK (33). An omnibus six marker test was carried out and we tested whether any SNP had an effect independent of the background haplotypic association.

Imputation analyses were carried out using Markov Chain Haplotyping software (MaCH 1.0) using phased data from CEU individuals from release 22 of the HapMap project (www.hapmap.org) as the reference set of haplotypes (www.sph.umich.edu/csg/abecasis/mach/index.htm). We only analyzed SNPs that were either genotyped or could be imputed with relatively high confidence ($r^2 > 0.3$). Association analysis of imputed SNPs was performed assuming an underlying additive model using the software package MACH2DAT (Li, Willer, Ding, Scheet and Abecasis, unpublished data) which accounts for uncertainty in prediction of the imputed data by weighting genotypes by their posterior probabilities. To evaluate imputed *P*-values against LD blocks we constructed an LD plot with the LDMAX and GOLD programs using the HapMap and genotype data used for imputation (32).

Comparison with erap1 *cis*-SNP expression data

A recent study of 400 children with/without childhood asthma generated strong evidence that certain *ERAP1* expression traits are both heritable and under *cis*-acting regulatory control. Dixon and colleagues genotyped 400 children from 206 families (recruited through a proband with childhood asthma) across a panel of >400 000 genome-wide SNPs. Subsequently, by microarray, they measured transcript abundance levels of 20 599 genes in EBV-transformed Lymphoblastoid cell lines (20). Finally, they tested for association between the genotype at each locus and the abundance of each transcript enabling the generation of a comprehensive expression QTL (eQTL) database. We searched this database for sites which influence the abundance of *ERAP1*. Data were available from a total of seven Affymetrix U133 Plus 2.0 probes, three of which (209788_s_at, 210385_s_at and 214012_at) detected strong *ERAP1* expression with maximal significance at marker rs28096 (P-value $= 1.6 \times 10^{-32}$) observed through probe 209788_s_at. Variation in heritability and expression between probes may indicate the presence of *ERAP1* splice variants, only a proportion of which are subject to genetic regulation (in particular at this site). Together these data suggest that a number of SNPs have significant *cis*-acting effect(s) on the expression of *ERAP1*, and that this effect (or effects) may be transcript specific. Therefore, we tested for a correlation between association of SNPs with AS from our study and association with *ERAP1* expression from their data.

Molecular modeling of ERAP1 structure

Human ERAP1 was modeled using the ICM program (Molsoft, San Diego) with the M1 aminopeptidase family

member tricorn-interacting factor 3 (34) (PDB id 1z1w), which shares 25.5% homology with ERAP1, as the structural template.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors would like to thank the Osteoarthritis Group, Oxford for the use of their control samples, Lyn-Louise Johnson for iPLEX genotyping at the WTCHG and acknowledge the assistance provided by WenHwa Lee for the modeling of ERAP1. We are grateful for additional financial support from the National Ankylosing Spondylitis Society and to their many colleagues in the UK for allowing them to study their patients. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from http://www.WTCCC.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

Conflict of Interest statement. None declared.

FUNDING

D.H. is funded by the NASS. T.K. is funded by the Henni Mester studentship. This study was funded, in part, by the Arthritis Research Campaign (UK), by the Wellcome Trust (076113) and by the National Institute for Health Research Oxford Biomedical Research Centre ankylosing spondylitis chronic disease cohort (Theme Code:A91202). The Structural Genomics Consortium is a registered charity (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. MAB is funded by a Principal Research Fellowship from the National Health and Medical Research Council (Australia).

REFERENCES

1. Laval, S.H., Timms, A., Edwards, S., Bradbury, L., Brophy, S., Milicic, A., Rubin, L., Siminovitch, K.A., Weeks, D.E., Calin, A. *et al.* (2001) Whole-genome screening in ankylosing spondylitis: evidence of non-MHC genetic-susceptibility loci. *Am. J. Hum. Genet.*, **68**, 918–926.
2. Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J. *et al.* (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.*, **39**, 1329–1337.
3. Tsui, F.W., Haroon, N., Reveille, J.D., Rahman, P., Chiu, B., Tsui, H.W. and Inman, R.D. (2009) Association of an ERAP1 ERAP2 haplotype with familial ankylosing spondylitis. *Ann Rheum Dis*. May 10 [Epub ahead of print], PMID: 19433412.
4. Maksymowych, W.P., Inman, R.D., Gladman, D.D., Reeve, J.P., Pope, A. and Rahman, P. (2009) Association of a specific ERAP1/ARTS1 haplotype with disease susceptibility in ankylosing spondylitis. *Arthritis Rheum.*, **60**, 1317–1323.
5. Cui, X., Hawari, F., Alsaaty, S., Lawrence, M., Combs, C.A., Geng, W., Rouhani, F.N., Miskinis, D. and Levine, S.J. (2002) Identification of ARTS-1 as a novel TNFR1-binding protein that promotes TNFR1 ectodomain shedding. *J. Clin. Invest.*, **110**, 515–526.
6. Adamik, B., Islam, A., Rouhani, F.N., Hawari, F.I., Zhang, J. and Levine, S.J. (2008) An association between RBMX, a heterogeneous nuclear ribonucleoprotein, and ARTS-1 regulates extracellular TNFR1 release. *Biochem. Biophys. Res. Commun.*, **371**, 505–509.
7. Islam, A., Adamik, B., Hawari, F.I., Ma, G., Rouhani, F.N., Zhang, J. and Levine, S.J. (2006) Extracellular TNFR1 release requires the calcium-dependent formation of a nucleobindin 2-ARTS-1 complex. *J. Biol. Chem.*, **281**, 6860–6873.
8. Schall, T.J., Lewis, M., Koller, K.J., Lee, A., Rice, G.C., Wong, G.H., Gatanaga, T., Granger, G.A., Lentz, R., Raab, H. *et al.* (1990) Molecular cloning and expression of a receptor for human tumor necrosis factor. *Cell*, **61**, 361–370.
9. Cui, X., Rouhani, F.N., Hawari, F. and Levine, S.J. (2003) Shedding of the type II IL-1 decoy receptor requires a multifunctional aminopeptidase, aminopeptidase regulator of TNF receptor type 1 shedding. *J. Immunol.*, **171**, 6814–6819.
10. Cui, X., Rouhani, F.N., Hawari, F. and Levine, S.J. (2003) An aminopeptidase, ARTS-1, is required for interleukin-6 receptor shedding. *J. Biol. Chem.*, **278**, 28677–28685.
11. Saric, T., Chang, S.C., Hattori, A., York, I.A., Markant, S., Rock, K.L., Tsujimoto, M. and Goldberg, A.L. (2002) An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nat. Immunol.*, **3**, 1169–1176.
12. Kisselev, A.F., Akopian, T.N., Woo, K.M. and Goldberg, A.L. (1999) The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J. Biol. Chem.*, **274**, 3363–3371.
13. Androlewicz, M.J., Anderson, K.S. and Cresswell, P. (1993) Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. *Proc. Natl. Acad. Sci. USA*, **90**, 9130–9134.
14. Neefjes, J.J., Momburg, F. and Hammerling, G.J. (1993) Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science*, **261**, 769–771.
15. Gromme, M. and Neefjes, J. (2002) Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways. *Mol. Immunol.*, **39**, 181–202.
16. van Endert, P.M., Tampe, R., Meyer, T.H., Tisch, R., Bach, J.F. and McDevitt, H.O. (1994) A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity*, **1**, 491–500.
17. York, I.A., Chang, S.C., Saric, T., Keys, J.A., Favreau, J.M., Goldberg, A.L. and Rock, K.L. (2002) The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat. Immunol.*, **3**, 1177–1184.
18. Chang, S.C., Momburg, F., Bhutani, N. and Goldberg, A.L. (2005) The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a 'molecular ruler' mechanism. *Proc. Natl. Acad. Sci. USA*, **102**, 17107–17112.
19. Hattori, A., Matsumoto, K., Mizutani, S. and Tsujimoto, M. (2001) Genomic organization of the human adipocyte-derived leucine aminopeptidase gene and its relationship to the placental leucine aminopeptidase/oxytocinase gene. *J. Biochem.*, **130**, 235–241.
20. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
21. Goto, Y., Hattori, A., Ishii, Y. and Tsujimoto, M. (2006) Reduced activity of the hypertension-associated Lys528Arg mutant of human adipocyte-derived leucine aminopeptidase (A-LAP)/ER-aminopeptidase-1. *FEBS Lett.*, **580**, 1833–1838.

22. Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.*, **3**, 85–98.
23. Orengo, J.P. and Cooper, T.A. (2007) Alternative splicing in disease. *Adv. Exp. Med. Biol.*, **623**, 212–223.
24. Pastinen, T., Ge, B., Gurd, S., Gaudin, T., Dore, C., Lemire, M., Lepage, P., Harmsen, E. and Hudson, T.J. (2005) Mapping common regulatory variants to human haplotypes. *Hum. Mol. Genet.*, **14**, 3963–3971.
25. Ouyang, C., Smith, D.D. and Krontiris, T.G. (2008) Evolutionary signatures of common human *cis*-regulatory haplotypes. *PLoS ONE*, **3**, e3362.
26. Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H. and Lee, M.P. (2003) Allelic variation in gene expression is common in the human genome. *Genome Res.*, **13**, 1855–1862.
27. Lesage, S., Zouali, H., Cezard, J.P., Colombel, J.F., Belaiche, J., Almer, S., Tysk, C., O'Morain, C., Gassull, M., Binder, V. *et al.* (2002) CARD15/NOD2 mutational analysis and genotype–phenotype correlation in 612 patients with inflammatory bowel disease. *Am. J. Hum. Genet.*, **70**, 845–857.
28. Nophar, Y., Kemper, O., Brakebusch, C., Englemann, H., Zwang, R., Aderka, D., Holtmann, H. and Wallach, D. (1990) Soluble forms of tumor necrosis factor receptors (TNF-Rs). The cDNA for the type I TNF-R, cloned using amino acid sequence data of its soluble form, encodes both the cell surface and a soluble form of the receptor. *Embo. J.*, **9**, 3269–3278.
29. Aderka, D., Engelmann, H., Wysenbeek, A.J. and Levo, Y. (1992) The possible role of tumor necrosis factor (TNF) and its natural inhibitors, the soluble-TNF receptors, in autoimmune diseases. *Isr. J. Med. Sci.*, **28**, 126–130.
30. Serwold, T., Gaw, S. and Shastri, N. (2001) ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. *Nat. Immunol.*, **2**, 644–651.
31. van der Linden, S., Valkenburg, H.A. and Cats, A. (1984) Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum.*, **27**, 361–368.
32. Excoffier, L. and Slatkin, M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.
33. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
34. Kyrieleis, O.J., Goettig, P., Kiefersauer, R., Huber, R. and Brandstetter, H. (2005) Crystal structures of the tricorn interacting factor F3 from Thermoplasma acidophilum, a zinc aminopeptidase in three different conformations. *J. Mol. Biol.*, **349**, 787–800.