Article type : Full Length

Genetic variants in *ERAP1* and *ERAP2* associated with immune-mediated diseases influence protein expression and isoform profile

Running title: *ERAP* expression in immune-mediated diseases

Aimee L. Hanson[1], Thomas Cuddihy[1], Katelin Haynes[1], Dorothy Loo[1], Craig J. Morton[2], Udo Oppermann[3], Paul Leo[1,4], Gethin P. Thomas[1,5], Kim-Anh Lê Cao[1], Tony J. Kenna*[1,4], Matthew A. Brown*[1,4]

Supported by Queensland Government Premier's Fellowship for Science and NHMRC Senior Principal Research Fellowship APP1024879 awarded to MAB.

[1] Aimee L. Hanson, BSc Hons, Thomas Cuddihy, MSc, Katelin Haynes, PhD, Dorothy Loo, MSc, Kim-Anh Lê Cao, PhD: Diamantina Institute, The University of Queensland, Translational Research Institute, Princess Alexandra Hospital, Queensland, Australia; [2] Craig J. Morton, PhD: ACRF Rational Drug Discovery Centre, St. Vincent's Institute of Medical Research, Fitzroy, Victoria, Australia; [3] Udo Oppermann, PhD: Botnar Research Center, Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, NIHR Oxford BRU, University of Oxford, United Kingdom; [4] Paul Leo, PhD, Tony J. Kenna, PhD, Matthew A. Brown, MBBS, MD, FRACP, FAHMS, FAA: Institute of Health and Biomedical Innovation, Queensland University of Technology, Translational Research Institute, Princess Alexandra Hospital, QLD, Australia; [5] Gethin P. Thomas, PhD: Charles Sturt University, Wagga Wagga, NSW, Australia

*Joint contribution by these senior authors

Address correspondence to:

Matthew A. Brown
Queensland University of Technology
Translational Research Institute
37 Kent Street
Woolloongabba
QLD 4102
Australia
Email: matt.brown@qut.edu.au
Phone: +61 7 3443 7364

ABSTRACT

Objective. The endoplasmic reticulum aminopeptidases ERAP1 and ERAP2, encoded on chromosome 5q15, trim endogenous peptides for human leukocyte antigen (HLA) mediated presentation to the immune system. Polymorphisms in *ERAP1* and/or *ERAP2* are strongly associated with several immune-mediated diseases with specific HLA backgrounds, implicating altered peptide handling and presentation as a prerequisite for autoreactivity against an arthritogenic peptide. Given the thorough characterisation of disease risk-associated polymorphisms that alter ERAP activity, this study aimed instead to interrogate the expression effect of chromosome 5q15 polymorphisms to determine their effect on ERAP isoform and protein expression. *Methods.* RNA sequencing and genotyping across chromosome 5q15 was used to detect genetic variants in *ERAP1* and *ERAP2* associated with altered total gene and isoform-specific expression. The functional implication of a putative mRNA splice-altering variant on ERAP1 protein levels was validated using mass spectrometry. *Results.* Polymorphisms associated with ankylosing spondylitis significantly influence the transcript and protein expression of the ERAP aminopeptidases. Disease risk-associated polymorphisms in and around both genes are also associated with increased gene expression. Furthermore, key risk-associated *ERAP1* variants are associated with altered transcript splicing, leading to allele-dependent alternate expression of two distinct isoforms, and significant differences in the type of ERAP1 protein produced. *Conclusions.* In accordance with studies demonstrating that polymorphisms that increase aminopeptidase activity predispose to immune disease, the elevated risk also attributed to increased expression of *ERAP1* and *ERAP2* supports the therapeutic notion of aminopeptidase inhibition to treat AS and other ERAP associated conditions.

Human leukocyte antigen (HLA)-mediated presentation of endogenous peptides to $CD8^+$ T-cells educates the immune system to differentiate between self-derived and foreign antigens. Consequently, aberrations in peptide presentation pathways may evoke misinformed immune cell reactivity against host tissue, leading to pathology. Aminopeptidases function in the processing of peptide precursors to generate molecules of optimal length and composition for HLA loading. Disease risk-associated polymorphisms in the endoplasmic reticulum aminopeptidase genes *ERAP1* and/or *ERAP2* are seen in several immune-mediated diseases, including ankylosing spondylitis (AS) (1, 2), psoriasis (3), Crohn's disease (CD) (4) and Behçet's disease (5).

Robust genetic interactions exist between *ERAP1* and the HLA class 1 allele *HLA-B27* and *HLA-B40* in AS (1, 6), *HLA-Cw6* in psoriasis (3), and *HLA-B51* in Behçet's disease (5). Synergism between HLA Class I loci and *ERAP1* is most pronounced in AS, with risk-associated polymorphisms in the gene contributing to strong disease risk (lead SNP rs30187 $P=4.4x10^{-45}$, OR=1.29;(7) in *HLA-B27* carriers alone (1). Whilst 80-95% of AS patients carry the major genetic risk factor *HLA-B27*, this allele contributes only ~20% of the genetic risk for the disease and over 113 further genetic loci have confirmed disease associations (1, 2, 7-10). Evidence points to the involvement of altered peptide handling in disease pathogenesis (11-13) potentially driving production of an immunogenic HLA-B27 specific peptide repertoire that misinforms the immune response and/or impedes the folding, stability and function of HLA-B27.

In AS the *ERAP1* locus contains two independent association signals (1). The strongest is tagged by SNP rs30187 (K528R; OR=1.29, $P=4.4x10^{-45}$;(7), the protective allele at which confers a ~40% decrease in enzymatic activity (1). Conversely, the rs30187 risk allele alters the B27-bound peptidome by causing both elevated production and destruction of different HLA-B27 epitopes (11, 14-16). The second association signal is tagged by rs10050860 (D575N; $P=9.28x10^{-36}$). This SNP has previously been shown not to be functional in both recombinant protein and cellular assays (1, 17), and the functional mechanism driving disease association at this signal has not been determined. *In vitro* studies have demonstrated that *ERAP1* protective variants, as well as *ERAP1* silencing or inhibition, reduces surface expression of HLA-B27 free heavy chains (18), although this finding has not been universal (19). HLA-B27 free heavy chain expression on the cell surface has been shown to induce the expansion of Th17 cells and secretion of the pro-inflammatory cytokine IL-17A (20). *ERAP1* homozygous deletion in the HLA-B27 transgenic rat model of AS does

not alter levels of folded HLA-B27 or HLA-B27 heavy chain dimers on the surface of peripheral blood mononuclear cells, and does not prevent the development of spondyloarthritis in these rats (21). The relevance of this to human spondyloarthritis however is unclear. Rats lack ERAP2 which is thought in humans to interact with ERAP1 to influence peptide presentation, and because the transgenic rats involved have upward of 20 copies of the HLA-B27 transgene per cell, as opposed to one or two copies in human HLA-B27 positive AS, they may develop spondyloarthritis as a result of a different pathogenic process to human disease. Evidence suggests that disease-associated polymorphisms in *ERAP1* also exert an expression effect (12), with risk genotype correlated with increased gene expression (22).

At *ERAP2* the AS-associated nonsense mutation, rs2248374, is an expression quantitative trait loci (eQTL) at which the protective G allele results in complete loss of gene expression due to nonsense-mediated decay of an alternatively spliced transcript (23). Presence of ERAP2 influences the HLA-B27 peptide pool, decreasing the abundance of peptides with N-terminal basic residues, and increasing the percentage of 9-mer peptides presented (24), lowering the affinity of the HLA-B27 peptidome (16, 25). Given the intricate role of both aminopeptidases in shaping the HLA-B27 peptidome, it is possible that variants increasing aminopeptidase expression may act concordantly to increase AS risk in individuals carrying disease-associated active variants of *ERAP1* and *ERAP2*.

In this study we used RNA-Seq to identify *ERAP1* and *ERAP2* polymorphisms associated with varied gene and isoform expression in AS cases and healthy controls (HCs). RNA-Seq allows interrogation of the transcriptional landscape of a gene, enabling identification of splice-altering variants (sQTLs) that may drive the production of distinct isoforms with phenotypic consequences. Despite extensive characterisation of alternate *ERAP2* transcripts that radically transform the global expression of this gene (23), alternate isoforms of *ERAP1* have been infrequently studied in the broader literature. We demonstrate that AS risk-associated SNPs in *ERAP1* and *ERAP2* are associated with substantial eQTL effects at both the transcript and protein level, and show that risk-associated variants in *ERAP1* are associated with altered splicing of the encoded gene transcript. This indicates that the disease-associated SNPs in *ERAP1* and *ERAP2* influence disease risk through effects on transcription and splice variation, in addition to the functional effects of coding variants in *ERAP1*.

PATIENTS AND METHODS

Ethics

Human ethics approval was granted by the Princess Alexandra Hospital and The University of Queensland Ethics Committees (ethic no. Metro South HREC/05/QPAH/221 and UQ 2006000102). Written informed consent was received from all participants prior to inclusion in the study.

Sample Selection

54 AS patients, diagnosed according to the modified New York criteria (26), and 70 HCs were included in this study (Table S5). Peripheral venous blood was collected from AS patients at the Princess Alexandra Hospital Ankylosing Spondylitis Clinic and peripheral blood mononuclear cells (PBMCs) from patients and controls extracted as described previously (7).

Genotyping, Imputation and Disease Association Analysis

Genotyping was conducted during the 2013 Immunochip study of 9049 AS cases and 13607 HCs (7). *P*-values for SNP disease associations upon conditioning with rs30187 were derived using logistic regression on the complete Immunochip data set in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/; (27), with inclusion of rs30187 genotype as a covariate. Imputation of SNPs within an R^2 window of 0.1 with rs30187 (*ERAP1*) or rs2910686 (*ERAP2*) was conducted using the 1000 Genomes Phase 1 reference panel for the subset of 124 individuals used in this study. Phasing was conducted using SHAPEIT (28) and imputation using IMPUTE2 (29) with the 'info' metric used to remove poorly imputed SNPs (info < 0.5). A total of 1221 SNPs spanning *ERAP1* and *ERAP2* were used in the final analysis.

RNA Sequencing

RNA was reverse transcribed, prepared for sequencing using Illumina TruSeq Standard Total RNA Library Prep Kit, and sequenced on an Illumina HiSeq 2000. Reads were mapped to the human genome Ensembl Reference Consortium build 37, release 75 (GRCh37.75) using TopHat version 2.0.6 employing the Bowtie 2 version 2.0.2 aligner (30, 31). Aligned reads were supplied to HTSeq (32) to generate counts per gene. Isoform-specific reads were generated using the Cufflinks suit (33) and assembled into a reference transcriptome to which RSEM (34) was used to align reads to generate isoform counts of known and novel transcripts. Gene and isoform counts were normalised using DESeq2 (35).

Statistical Analysis for eQTL Detection

All statistical analyses were performed using the R statistical software (36). A generalised linear mixed model (GLMM) with a negative binomial distribution and logarithmic link was fitted using the lme4 package (37) to test for the effect of genotype on gene or isoform expression, correcting for the fixed effect of patient sex and random effect of sequencing batch:

```
glmer(expression ~ genotype + sex + (1|batch. no.), family = neg.bin)
```

The change in gene or isoform expression attributed to the addition of a minor allele was quantified as the exponential of the genotype coefficient returned from the model (referred to as the incident rate ratio, IRR), with IRR^2 comparable to the fold change in expression between individuals with minor allele count 2 over count 0. *P*-values were adjusted for multiple testing using Benjamini and Hochberg false discovery rate correction (38). Significant SNPs were cross-referenced to the set of significant disease-associated SNPs previously identified at this locus (7), and the direction of effect of the disease risk allele inferred. Pairwise conditioning was conducted by adding each significant eQTL individually as a covariate into the GLMM and assessing the genotype effect on expression at each remaining SNP:

```
glmer(expression ~ genotype + sex + conditional SNP genotype + (1|batch.
                                    no.), family = neg.bin)
```

A Wilcoxon test was used to test for differences in gene expression between cases and controls.

Statistical Analysis for sQTL Detection

Splice-altering quantitative trait loci (sQTLs) were found for two highly expressed isoforms of *ERAP1*, herein termed isoform 19E and 20E. Isoform 19E proportion was calculated from isoform counts as isoform 19E / (isoform 19E + 20E) for each individual. A two-sample t-test was used to assess the difference in isoform proportion between cases and controls. A linear model was applied to test for the effect of genotype on isoform proportion while correcting for any disease status effect:

```
lm(isoform 19E proportion ~ genotype + status)
```

SNPs that were determined to be significant eQTLs for both transcripts, and significant sQTLs with an effect on isoform proportion, were cross-referenced to the set of significant disease-associated SNPs previously identified at this locus (7), and the direction of the effect

of the disease risk allele inferred. A Wilcoxon test was used to test for differences in isoform expression between cases and controls.

Statistical Analysis of Haplotype Data

A chi-squared test assessed the disease association of a haplotype containing the ERAP1 activity altering SNP rs30187 and splice-altering SNP rs7063 using the European cohort of 9,069 AS cases and 13,578 HCs genotyped in the Immunochip investigation (7).

Mass Spectrometry for Quantification of ERAP1 Isoform and Total Protein Expression

PBMCs from 39 samples (cases and controls; 15 homozygous for the risk (major) allele at rs7063, 15 heterozygous and 9 homozygous for the protective allele; Table S5) were prepared for mass spectrometry as per the Supplementary Methods. The following peptides were targeted for detection and quantification in samples using the Agilent 1260 Chip HPLC-6490 triple quadrupole mass spectrometer (Table S6 for details):

ERAP1 isoform 19E: VW**L**QSEKLER

ERAP1 isoform 20E: VWLQSEK**L**EHDPEADATG

ERAP1 common peptide: NP**V**GYPLAWQFLR

Samples were spiked with 0.5fmol weighted peptide mix standard for quantification, and 8µL (~2µg) was subjected to mass spectrometry. Standard curve was generated by injecting 1µL peptide mix standards at 0.1fmol/µ 0.2fmol/µL, 0.4fmol/µL, 0.6fmol/µL, 0.8fmol/µL.

Statistical Analysis of Mass Spectrometry Data

Mass chromatograms were analysed using Skyline (39) to determine ERAP1 19E, 20E and total protein concentrations in each sample. A linear model was used to test for the effect of rs7063 genotype on protein expression for each isoform and total protein, correcting for the effect of disease status and *HLA-B27* status. A *t*-test was used to test for differences in isoform or total ERAP1 expression between cases and controls.

Protein modelling of ERAP1 Isoform 20E

The sequence of human ERAP1 isoform 20E was taken from the Uniprot sequence entry Q9NZ08-2. The sequence was submitted to the online modelling server Phyre2 (40) in intensive mode. A set of four templates were used by the server to generate the final model, with 92% of the sequence modelled at >90% confidence. A total of 80 residues (residues 1-45, 487-513 and 941-948) were built *ab initio,* including residues in the C-terminal extension

caused by the additional exon 20.

RESULTS

Disease risk SNPs in *ERAP1* and *ERAP2* are associated with increased total gene expression. Total gene expression information (normalised transcript counts) was tested for an association with genotype at 1,221 genotyped and imputed SNPs across chromosome 5q15. There were 113 disease risk-associated SNPs identified as eQTLs significantly influencing total *ERAP1* expression (Table S1). At every SNP the disease risk allele was associated with an increase in gene expression over the corresponding protective allele. Pairwise conditioning identified 38 SNPs as variants controlling the expression effect at this locus (Table S1); conditioning with any of these 38 abolished the significance of all 113 original eQTLs. The 38 variants span a region from between the 19th intron of the gene to approximately 50kb upstream of the first exon (Fig.1A). All 38 SNPs lost their disease association but retained their expression association upon conditioning on rs30187, which itself displayed a modest eQTL effect (IRR^2=1.185, P=2.4x10^{-3}). Imputed SNP rs39840 (linked with genotyped SNPs rs27038 and rs27041, AS disease associations P=5.7x10^{-19} and P=4.9x10^{-19} respectively) was the most significant eQTL identified (IRR^2=1.343, P=1.5x10^{-7}), at which risk homozygous genotype conferred a 34.3% increase in *ERAP1* expression over the protective homozygous genotype (Fig.1B, Table S1). There was no significant difference in mean *ERAP1* expression between AS cases and HCs.

ERAP2 expression has a bimodal distribution due to the effects of the rs2248374 nonsense mutation found at approximately 50% frequency in the population (23). To generating an expression distribution amendable to modelling, 38 rs2248374-GG individuals, expressing very low to no *ERAP2* transcript, were removed, and the rs2248374 genotype was corrected for to ablate the expression effect of this SNP in heterozygotes. There were 156 SNPs that exhibited a significant association with *ERAP2* expression. Pairwise conditioning identified 94 SNPs that controlled for the *ERAP2* expression association of all others (Table S2). Tight linkage disequilibrium between SNPs across the entirety of the *ERAP2* locus (Fig.1C) made it difficult to pinpoint a position for the eQTL signal. As at *ERAP1*, disease-risk genotypes were consistently associated with increased gene expression at all expression-controlling SNPs within and about the *ERAP2* gene, with the risk genotype at the top eQTLs resulting in a 148% elevation in expression (IRR^2=2.476, P=1.7x10^{-6}; Fig.1D). There was no significant difference in mean *ERAP2* expression between AS cases and HCs; this study had

low power to detect such a difference given the small sample size (38 cases and 48 controls) upon removal of rs2248374 homozygotes, and the need to control for rs2248374 (which is disease associated) in the analysis.

Alternate expression of two *ERAP1* isoforms is governed by genotype and associated with disease. Eleven unique *ERAP1* isoforms were assembled from the RNA-Seq data (Fig.S1A). Only two transcripts, differing in the inclusion of the C-terminal exon 20 (RM[STOP] → HDPEADATG[STOP]), Ensembl ID ENST00000443439 (941 amino acids, 19 exons; 19E) and ENST00000296754 (948 amino acids, 20 exons; 20E) (Fig.2, Table S3), were highly expressed in all individuals (Fig.S1B). We identified 158 disease-associated SNPs as significant eQTLs for both isoforms when tested independently for an expression effect on each (Table S4). At every SNP, the disease risk variant was associated with a significant increase in the expression of the isoform 19E and a significant decrease in the expression of the isoform 20E. The proportion of isoform 19E transcript expressed by an individual was significantly associated with disease status (P=0.047, T-test), with cases expressing a significantly greater proportion of the 19-exon isoform (66.1% 19E) than controls (61.9% 19E).

Genotype at the same 158 disease-associated SNPs (henceforth called sQTLs) had significant effects on *ERAP1* isoform proportions. Pairwise conditioning identified 9 SNPs that controlled for the sQTL effect; conditioning with any of these 9 ablated the effect of all other SNPs on the expression of isoform 20E, and thus their effect on isoform proportion (Table S4). The SNP rs7063 (disease association OR=1.34, P=1.3x10^{-41};(7), situated between exon 19 and 20 (Fig.2), exhibited the most significant effect on isoform proportion (OR=1.16, P=1.2x10^{-23}), at which risk (major allele) homozygotes expressed 105% more (IRR^2=2.05, P=8.7x10^{-12}) and 47% less (IRR^2=0.53, P=1.0x10^{-17}) of the 19-exon and 20-exon isoforms respectively than protective homozygotes (Fig.3A, 3B). rs7063 heterozygotes expressed similar amounts of both transcripts (56% isoform 19E, 44% isoform 20E), whereas risk allele homozygotes expressed predominantly isoform 19E (71% on average) and protective allele homozygotes predominantly isoform 20E (59% on average; Fig.3D). None of the 58 most significant sQTLs exhibited significant eQTL effects on total *ERAP1* expression, and all retained a significant disease association upon correction for rs30187 (Table S4) but lost this association upon correction for both rs30187 and rs10050860. These SNPs clustered within and about the C-terminus encoding exons 19 and 20 (Fig.3C). Upon correction for rs10050860 alone, rs7063 retained a genome-wide significant disease

association (P=1.4x10^{-10}). Conversely, a far weaker association remained for rs10050860 (P=2.6x10^{-5}) following correction for rs7063. LD between rs7063 and rs10050860 (R^2=0.55) is considerably stronger than between rs7063 and rs30187 (R^2=0.21).

A strong disease-protective haplotype results from co-occurrence of the protective alleles at rs30187 and the sQTL rs7063. The protective haplotype, rs30187-C/rs7063-T (low ERAP1 enzyme activity/decreased expression of *ERAP1* isoform 19E, increased expression of isoform 20E), showed a disease association of P=1.1x10^{-61} (Table 1) in 9069 genotyped AS cases and 13578 HCs. This haplotype association exceeded the individual disease association of both SNPs in isolation (rs30187: P=4.4x10^{-45}; rs7063: P=1.3x10^{-41}).

Genotype at *ERAP1* sQTL rs7063 is strongly associated with alternate expression of two ERAP1 isoforms at the protein level. Protein derived from both the 19E and 20E isoforms of the *ERAP1* transcript was detected using mass spectrometry. SNP genotype at rs7063 was significantly associated with protein expression of both isoforms independently, and total ERAP1 expression (the sum of the expression of both isoforms; Fig.3F and 3G). Individuals homozygous for the risk allele at rs7063 expressed 3.52 times more isoform 19E (P=3.1x10^{-6}) and 0.29 times less isoform 20E (P=2.1x10^{-5}), and expressed 2.37 times more total ERAP1 protein overall (P=5.6x10^{-5}) than those homozygous for the protective allele. At the protein level, isoform 19E was the predominantly expressed form of ERAP1 in all three genotype groups (Fig.3E), expressed significantly higher on average than isoform 20E (P=1.4x10^{-10}). There was no significant difference in the mean expression of either isoform at the protein level, or total ERAP1 protein levels, between AS cases and HCs.

***In silico* analysis of *ERAP1* isoform 20E predicts folding nature of the alternate protein.** *Ab initio* protein structure modelling of the ERAP1 20E isoform suggests that the unique additional residues on the C-terminal end of the protein fold back across the surface of domain 4 (D4) of ERAP1, potentially forming salt-bridge interactions with residues Arg750 and Arg708 (Fig.4).

DISCUSSION

The genetic association of variants in *ERAP1* and *ERAP2* with several immune-mediated diseases make these enzymes important targets for functional investigation. This study showed that AS risk-associated variants in these genes consistently demonstrate an eQTL

effect that increases aminopeptidase expression. Indeed, the protective influence of the nonsense mutation rs2248374 (23) in *ERAP2* implies that some aspect of the enzyme's functional role in peptide presentation is linked to pathology in a way that can be ablated with loss of expression. It is expected that this is also the case at *ERAP1*, at which increased expression may be exacerbating the pathogenic effect of co-occurring missense mutations shown to alter enzymatic activity and peptide handling (11, 12). It has been previously shown that *ERAP1* mRNA and protein expression is elevated in lymphoid cell lines carrying disease susceptibility variants across *ERAP1* haplotypes (15, 22). Our results validate the influence of genetic variants across *ERAP1* and *ERAP2* on transcript and protein expression levels.

Previous fine-mapping and haplotype evolution studies indicate that association at *ERAP1* with AS is driven by two independent association signals, one tagged by the coding variant rs30187 (K528R) which is likely the causative SNP, and the second tagged by the coding variant rs10050860 (D575N) (1, 6). Further studies using recombinant ERAP1 and cell lines carrying ERAP1 variants have shown that rs30187 has a substantial effect on peptide cleavage and peptides presented by HLA Class 1 antigens (1, 17, 41). A suggestion has been raised that *ERAP1* variants act in haplotypic combinations (42, 43). However, many of the variants and haplotypes reported in these studies have been shown either not to exist or to be found in markedly different frequencies to those reported in the literature that they likely represent experimental artefact (44, 45). In the current study we have aimed to investigate the functional mechanisms by which *ERAP1* and *ERAP2* variants operate, to explain the known genetic association of these genes with AS and other diseases.

The major finding of this study suggests complex regulation of *ERAP1* expression by two independently acting variants; one in moderate LD with rs30187 that influence global *ERAP1* expression, and the second in moderate LD with rs10050860 that influences alternate splicing of two distinct transcripts. Our results suggest that the differences between these two forms of the *ERAP1* transcript, and encoded protein, are critically important, and influence the levels of functional enzyme in cells. Given the cross-disease concordant action of *ERAP1* and *ERAP2* polymorphisms in conditions such as psoriasis and Crohn's disease (46), these finding have relevance far beyond AS alone.

Harvey *et al.* (12) demonstrated a strong positive correlation between the strength of AS disease association for variants in *ERAP1* and their effect on *ERAP1* expression (r=0.75) (47). Similarly, we noted that the most significant *ERAP1* eQTLs are strongly associated with AS. Although linked within the haplotype containing the key functional variant, rs30187, these SNPs retained a significant expression effect upon rs30187 correction. This implies

that, additional to the effects of rs30187 on ERAP1 activity, there are one or more variants in moderate LD with rs30187 driving altered enzyme expression. Joint inheritance of these alleles would mean that any alterations in the HLA-B27 peptidome due to ERAP1 hyperactivity would be exacerbated by enzyme overexpression. This could potentially flood HLA-B27 with an immunogenic repertoire stimulating CD8-mediated immune response (48), or lead to destruction of peptides that protect in some way against the immunological processes that lead to AS (49). Concordantly, increased transcript expression in individuals expressing *ERAP2* and carrying *ERAP2* disease risk-associated polymorphisms supports the significance of increased aminopeptidase expression in disease pathology. The identification of disease-associated expression-influencing variants at this locus, independent of the rs2248374 null mutant, implies that the portion of the population that do express ERAP2 have varied expression phenotypes and thus disease susceptibility.

Of great interest was the identification that two expressed isoforms of *ERAP1*, the alternate expression of which correlated with genotype at strong disease risk-associated SNPs, may play a substantial role in disease. These two transcripts, encoding different versions of the ERAP1 protein, have been previously identified (50, 51) and annotated as ENST00000443439 (19 exons, 18 coding; 19E) and ENST00000296754 (20 exons, 19 coding; 20E) in the Ensembl database (52). Isoform quantification on the transcript level revealed that 19E and 20E are both predominant forms *ERAP1*, expressed in all 124 individuals, and at high enough levels to potentially contribute to the functional output of the enzyme. We demonstrate here that alternate expression of these two isoforms is likely modulated by a genetic splice-interfering variant rather than by the common mechanism of alternate splicing, and that this variant is strongly AS associated. The observation that the disease risk-associated allele at all significant *ERAP1* sQTLs correlated with increased expression of the 19E transcript and decreased expression of the 20E transcript suggests there may be some disease protective feature of the 20E isoform specifically. SNPs exhibiting the most significant sQTL effect were localised around the exon/intron 19 boundary, around the first point of sequence variation between the two isoforms, and it has been previously noted that a number of the strongest disease-associated variants in *ERAP1* fall in this location with a potential involvement in splicing (12). The most significant sQTLs retained a strong disease association upon correction for rs30187, implying they are situated within the rs10050860 tagged haplotype; perhaps a splice site variant governing *ERAP1* isoform expression is of greatest functional importance at this position. Our data suggests that the rs10050860 association is driven by the sQTL rs7063, but we cannot formally exclude effects due to other

polymorphisms in strong linkage disequilibrium with it, noting that rs10050860 itself has been shown not to affect ERAP1 peptide cleavage (1, 17).

Mass spectrometric quantification of isoform expression showed that both the 19E and 20E forms of the ERAP1 protein were present at detectable levels in all 39 assayed samples. Likewise, observed differences in the molecular mass of ERAP1 isolated from various human cell lines confirms the co-occurrence of the two forms of the full-length protein (53). One putative mRNA splice-altering mutation (rs7063) in *ERAP1* contributed to marked variability in the type of ERAP1 protein expressed. This variant has an AS association of $P=1.3x10^{-41}$ (7) and $2.5x10^{-17}$ upon correction for rs30187. The observation that a far stronger association of rs7063 with AS remains after correction for rs10050860 ($P=1.4x10^{-10}$), than the converse (rs10050860 association having corrected for rs7063, $P=2.6x10^{-5}$) supports rs7063, or a marker in tight linkage disequilibrium with it, being the more important variant the second *ERAP1* disease-associated haplotype. rs7063 falls in the middle of the conserved transcription termination sequence (AA**T**AAA) in the 3′UTR of isoform 19E, a motif recognised by cleavage and polyadenylation specific factors involved in 3′-end transcription termination in mammals (54). The risk allele (major allele T on the reverse strand in the motif AA**T**AAA, allele A on the forward strand) would be expected to promote correct termination of the 19-exon form of the transcript, whereas the protective allele may result in loss of termination sequence recognition, producing the 20-exon form of the transcript. Variant rs111774449, falling in the splice donor sequence at the 3′ end of exon 19 (55), may also be a candidate for isoform switching but is rare (MAF < 0.01 ExAC) and was not genotyped or imputed in this study. The strong disease-protective haplotype that arises from co-occurrence of rs30187 and rs7063 protective alleles is evidence that *ERAP1* expression dynamics contribute to the pathogenicity afforded by increased enzyme activity. The enhanced genetic contribution of SNP haplotypes at 5q15 has been previously noted, with carriage of both the two disease-associated haplotypes increasing risk by ~4 fold, a far greater risk than the additive effects of either variant alone (1.2-1.3 fold for either variant in isolation) (1). The current study provides a potential functional mechanism for this observation.

Quantification of ERAP1 protein isoforms in individuals with different rs7063 genotypes implied differences in the dynamics of translation from mRNA to protein. Despite approximately equal levels of the 19E and 20E transcripts in rs7063 heterozygotes, 81% of the ERAP1 protein detected in these individuals was of the 19E form. This raises the question, is loss of isoform 20E post transcription due to sequence or structural variation

between the two isoforms (gain of a C-terminal exon), perhaps contributing to less efficient translation of isoform 20E, or misfolding and degradation of the protein, with subsequent lack of functionality? If this is the case, the mechanism of protection conferred by the disease protective genotype at splice site SNPs may arise due to skewed expression towards the 20E transcript, subsequent loss of this isoform at the protein level, and thus a decrease in the overall level of functional ERAP1 available to the cell.

Ab initio modelling of the additional amino-acid residues of isofom 20E was required to explore the role of the exon 20 residues in the structure of ERAP1. The area of the ERAP1 domain 4 with which the exon-20 derived residues are predicted to interact (Fig.4) is effectively invariant structurally between the open and closed forms of ERAP1, implying that the presence of the isoform 20E C-terminal extension is unlikely to alter the open to closed form dynamics believed to be required for substrate binding and release (56, 57). Tertiary structure modelling, however, could not predict whether these residues interfere with folding of the mature protein. The UCSC 100 vertebrates conservation track (58) shows very low sequence conservation in exon 20 relative to the other 19 exons of *ERAP1*, and, indeed, other coding exons in general. This suggests that the 19-exon ERAP1 is the predominant functional form of the enzyme and that the appended seven amino acids in isoform 20E are either irrelevant to the mature enzyme which is thus tolerant of substituting mutations, or that the protein itself is removed from the cell before it can actively contribute to peptide trimming.

To date, differential gene expression studies in AS have found no evidence for significant changes in the expression of genes encoding disease-associated aminopeptidases in patients relative to healthy controls (59, 60). Here we demonstrate the ability of eQTL studies to identify genotype-driven altered gene expression in disease, which would go unnoticed by differential expression studies that pool samples of different genotypes. Foremost, it must be acknowledged that dynamic changes in the isoforms derived from a transcriptional unit can be far more insightful than measures of total gene expression, adding layers of complexity to the functional output of a gene. It appears that increased aminopeptidase expression governed by altered transcript dynamics at both *ERAP1* and *ERAP2* is a key mechanism driving the degree to which these enzymes contribute to immune-mediated disease. These findings support ERAP inhibition as a therapeutic approach for treating diseases including ankylosing spondylitis and psoriasis, where protective genetic variants lead to loss of both ERAP levels and function.

DATA ACCESS

The RNA sequencing data used for transcript expression analysis have been submitted to NCBI's Sequence Read Archive (project no. SRP100652, experiment no. SRX2586069) with all sample metadata included within BioProject PRJNA376610.

ACKNOWLEDGEMENTS

MAB was supported by NHMRC Senior Principal Research Fellowship APP1024879. Funding for this project was provided by Queensland Government Premier's Fellowship for Science awarded to MAB. UO was supported by ARUK program grant 20522. The authors thank L.Bradbury and K.Hollis for their assistance in ascertainment of participant samples and all participants who provided samples and clinical details for this project. Thanks to B.Gautier for assistance in statistical model design and testing.

REFERENCES

1. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011;43(8):761-7.

2. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium, Burton PR, Clayton DG, Cardon LR, Craddock N, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet. 2007;39(11):1329-37.

3. Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010;42(11):985-90.

4. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010;42(12):1118-25.

5. Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, et al. Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet. 2013;45(2):202-7.

6. Cortes A, Pulit SL, Leo PJ, Pointon JJ, Robinson PC, Weisman MH, et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nature communications. 2015;6(7146).

7. International Genetics of Ankylosing Spondylitis Consortium, Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk variants for

ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet. 2013;45(7):730-8.

8. Australo-Anglo-American Spondyloarthritis Consortium, Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet. 2010;42(2):123-7.

9. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet. 2016;48:510-8.

10. Danoy P, Pryce K, Hadler J, Bradbury LA, Farrar C, Pointon J, et al. Association of variants at 1q32 and STAT3 with ankylosing spondylitis suggests genetic overlap with Crohn's disease. PLoS genetics. 2010;6(12):e1001195.

11. Seregin SS, Rastall DP, Evnouchidou I, Aylsworth CF, Quiroga D, Kamal RP, et al. Endoplasmic reticulum aminopeptidase-1 alleles associated with increased risk of ankylosing spondylitis reduce HLA-B27 mediated presentation of multiple antigens. Autoimmunity. 2013;46(8):497-508.

12. Harvey D, Pointon JJ, Evans DM, Karaderi T, Farrar C, Appleton LH, et al. Investigating the genetic association between ERAP1 and ankylosing spondylitis. Hum Mol Genet. 2009;18(21):4204-12.

13. Benjamin RJ, Abrams JR, Parnes JR, Madrigal JA, Parham P. Polymorphic specificity of Q1/28, a monoclonal antibody that preferentially reacts with free class I heavy chains. Immunogenetics. 1992;37(1):73-6.

14. Evnouchidou I, Birtley J, Seregin S, Papakyriakou A, Zervoudi E, Samiotaki M, et al. A common single nucleotide polymorphism in endoplasmic reticulum aminopeptidase 2 induces a specificity switch that leads to altered antigen processing. J Immunol. 2012;189(5):2383-92.

15. Sanz-Bravo A, Campos J, Mazariegos MS, Lopez de Castro JA. Dominant role of the ERAP1 polymorphism R528K in shaping the HLA-B27 peptidome through differential processing determined by multiple peptide residues. Arthritis & rheumatology. 2015;67(3):692-701.

16. Martin-Esteban A, Sanz-Bravo A, Guasp P, Barnea E, Admon A, Lopez de Castro JA. Separate effects of the ankylosing spondylitis associated ERAP1 and ERAP2 aminopeptidases determine the influence of their combined phenotype on the HLA-B*27 peptidome. J Autoimmun. 2017.

17. Garcia-Medel N, Sanz-Bravo A, Van Nguyen D, Galocha B, Gomez-Molina P, Martin-Esteban A, et al. Functional interaction of the ankylosing spondylitis-associated endoplasmic reticulum aminopeptidase 1 polymorphism and HLA-B27 in vivo. Mol Cell Proteomics. 2012;11(11):1416-29.

18. Chen L, Ridley A, Hammitzsch A, Al-Mossawi MH, Bunting H, Georgiadis D, et al. Silencing or inhibition of endoplasmic reticulum aminopeptidase 1 (ERAP1) suppresses free heavy chain expression and Th17 responses in ankylosing spondylitis. Ann Rheum Dis. 2016;75(5):916-23.

19. Haroon N, Tsui FW, Uchanska-Ziegler B, Ziegler A, Inman RD. Endoplasmic reticulum aminopeptidase 1 (ERAP1) exhibits functionally significant interaction with HLA-B27 and relates to subtype specificity in ankylosing spondylitis. Ann Rheum Dis. 2012;71(4):589-95.

20. Bowness P, Ridley A, Shaw J, Chan AT, Wong-Baeza I, Fleming M, et al. Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis. J Immunol. 2011;186(4):2672-80.

21. Barnea E, Melamed Kadosh D, Haimovich Y, Satumtira N, Dorris ML, Nguyen MT, et al. The Human Leukocyte Antigen (HLA)-B27 Peptidome in Vivo, in Spondyloarthritis-susceptible HLA-B27 Transgenic Rats and the Effect of Erap1 Deletion. Mol Cell Proteomics. 2017;16(4):642-62.

22. Costantino F, Talpin A, Evnouchidou I, Kadi A, Leboime A, Said-Nahal R, et al. ERAP1 gene expression is influenced by non-synonymous polymorphisms associated with predisposition to spondyloarthritis. Arthritis & rheumatology. 2015.

23. Andres AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin SQ, Hurle B, et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. PLoS genetics. 2010;6(10):e1001157.

24. Martin-Esteban A, Guasp P, Barnea E, Admon A, Lopez de Castro JA. Functional interaction of the ankylosing spondylitis associated endoplasmic reticulum aminopeptidase 2 with the HLA-B*27 peptidome in human cells. Arthritis and rheumatism. 2016;68(10):2468-75.

25. Garcia-Medel N, Sanz-Bravo A, Alvarez-Navarro C, Gomez-Molina P, Barnea E, Marcilla M, et al. Peptide handling by HLA-B27 subtypes influences their biological behavior, association with ankylosing spondylitis and susceptibility to endoplasmic reticulum aminopeptidase 1 (ERAP1). Mol Cell Proteomics. 2014;13(12):3367-80.

26. van der Linden S, Valkenberg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. Arthritis Rheum. 1984;27(4):361-8.

27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

28. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature methods. 2012;9(2):179-81.

29. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics. 2009;5(6):e1000529.

30. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.

31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

32. Anders S, Pyl PT, Huber W. HTSeq - A Python framework to work with high-throughput sequencing data. bioRxiv. 2014.

33. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-5.

34. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv. 2014.

36. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

37. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1.-8. 2015.

38. Beniamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Satistical Society. 1995;57(1):289-300.

39. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966-8.

40. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845-58.
41. Chen L, Fischer R, Peng Y, Reeves E, McHugh K, Ternette N, et al. Critical Role of Endoplasmic Reticulum Aminopeptidase 1 in Determining the Length and Sequence of Peptides Bound and Presented by HLA-B27. Arthritis & rheumatology. 2014;66(2):284-94.
42. Reeves E, Colebatch-Bourn A, Elliott T, Edwards CJ, James E. Functionally distinct ERAP1 allotype combinations distinguish individuals with Ankylosing Spondylitis. Proc Natl Acad Sci U S A. 2014;111(49):17594-9.
43. Reeves E, Elliott T, Edwards CJ, James E. Both rare and common ERAP1 allotypes have distinct functionality defined by polymorphic context and are important in AS association. Proc Natl Acad Sci U S A. 2017;114(9):E1575-E6.
44. Roberts AR, Appleton LH, Cortes A, Vecellio M, Lau J, Watts L, et al. ERAP1 association with ankylosing spondylitis is attributable to common genotypes rather than rare haplotype combinations. Proc Natl Acad Sci U S A. 2017;114(3):558-61.
45. Robinson PC, Brown MA. ERAP1 biology and assessment in Ankylosing Spondylitis. Proceedings of the National Academy of Sciences. 2015;112(15):E1816-E.
46. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013;14(9):661-73.
47. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide association study of global gene expression. Nat Genet. 2007;39(10):1202-7.
48. Benjamin R, Parham P. HLA-B27 and disease: a consequence of inadvertent antigen presentation? Rheum Dis Clin North Am. 1992;18(1):11-21.
49. Kenna TJ, Brown MA. Immunopathogenesis of ankylosing spondylitis. Int J Clin Rheumatol. 2013;8(2):265-74.
50. Hattori A, Matsumoto H, Mizutani S, Tsujimoto M. Molecular Cloning of A-LAP highly related to placental leucine aminopeptidase/oxytocinase. J Biochem. 1999;125:931-8.
51. Hattori A, Matsumoto K, Mizutani S, Tsujimoto M. Genomic Organisation of A-LAP gene and its relationship to the placental leucine aminopeptidease:ozytocinase. J Biochem. 2001;130(2):235-41.
52. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res. 2016;44(D1):D710-6.
53. Yousaf N, Low WY, Onipinla A, Mein C, Caulfield M, Munroe PB, et al. Differences between disease-associated endoplasmic reticulum aminopeptidase 1 (ERAP1) isoforms in

cellular expression, interactions with tumour necrosis factor receptor 1 (TNF-R1) and regulation by cytokines. Clin Exp Immunol. 2015;180(2):289-304.

54. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes Dev. 2009;23(11):1247-69.

55. National Center for Biotechnoligy Information. NCBI Human Genome Browser. U.S National Library of Medicine; 2015.

56. Kochan G, Krojer T, Harvey D, Fischer R, Chen L, Vollmar M, et al. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. Proc Natl Acad Sci U S A. 2011;108(19):7745-50.

57. Nguyen TT, Chang SC, Evnouchidou I, York IA, Zikos C, Rock KL, et al. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. Nat Struct Mol Biol. 2011;18(5):604-13.

58. University of California Santa Cruz (UCSC). UCSC Genome Browser: Vertebrate Multiz Alignment & Conservation (100 Species). 2015.

59. Duan R, Leo P, Bradbury L, Brown MA, Thomas G. Gene expression profiling reveals a downregulation in immune-associated genes in patients with AS. Ann Rheum Dis. 2010;69(9):1724-9.

60. Assassi S, Reveille JD, Arnett FC, Weisman MH, Ward MM, Agarwal SK, et al. Whole-blood gene expression profiling in ankylosing spondylitis shows upregulation of toll-like receptor 4 and 5. J Rheumatol. 2011;38(1):87-98.

TABLES

Table 1: Results of a Chi-Squared test for association of the haplotype rs7063, risk allele A; rs30187, risk allele T with ankylosing spondylitis in 6096 cases and 13578 HCs.

rs7063	rs30187	Haplotype Frequency	Frequency Controls	Frequency Cases	*P*-value
T	T	0.0458	0.0429	0.0501	3.2×10^{-4}

T	C	0.2360	0.2630	0.1955	1.1×10^{-61}
A	T	0.3194	0.2956	0.3552	1.7×10^{-40}
A	C	0.3988	0.3985	0.3992	0.89

FIGURE LEGENDS

Fig.1: Genotypes across chromosome 5q15 locus SNPs correlate with *ERAP1* and *ERAP2* gene expression. Locus plot of *P*-values for the *ERAP1* **(A)** and *ERAP2* **(C)** expression effect of chromosome 5q15 locus SNPs plotted against genomic region, with eQTLs controlling for the expression effect bracketed in red. Lead eQTLs are marked in purple and SNPs are coloured according to LD with these variant as per the colour key at the top of the plot. Box plot of lead *ERAP1* eQTL rs39840 (risk allele G; **B)** and lead

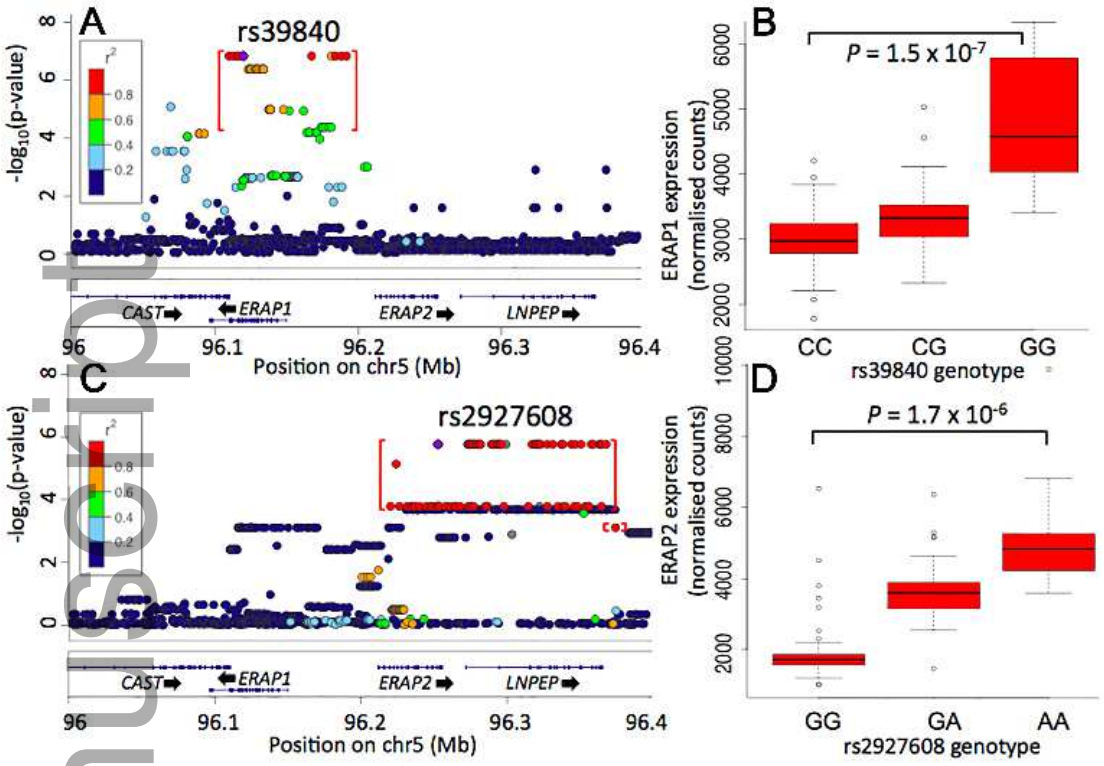
representative *ERAP2* SNP rs2927608 (risk allele A; **D)** genotype verses total transcript expression of the corresponding gene; *P*-values derived from a generalised linear mixed model regressing genotype against gene expression taken from normlised RNA-Seq counts. n = 79(CC), 39(CG) and 6(GG) for genotype groups in B and n = 44(GG), 54(GA) and 26(AA) for genotype groups in D.

Fig.2: Graphical representation (not to scale) of the 3′ end of two *ERAP1* transcripts found to be highly expressed in PBMCs. Isoform ENST00000443439 has 19 exons (19E) and isoform ENST00000296754 20 exons (20E) with an alternate 3′UTR sequence. Encoded amino acids are indicated above exons 19 and 20 (black boxes, numbered at top of figure) and encoding codons beneath, beginning two amino acids prior to the first point of variation between the transcripts (RM*→HDPEADATG*). The codon for amino acid H in isoform 20E spans a splice junction. Arrows indicate the location of two putative splice interfering SNPs; rs111774449 (G/A; Arg/His) at the exon-intron 19 interface, and rs7063 (A/T) falling within the transcription termination motif in the 3′UTR of isoform 19E, identified as the most significant *ERAP1* sQTL from the statistical analyses conducted on RNA-Seq data.

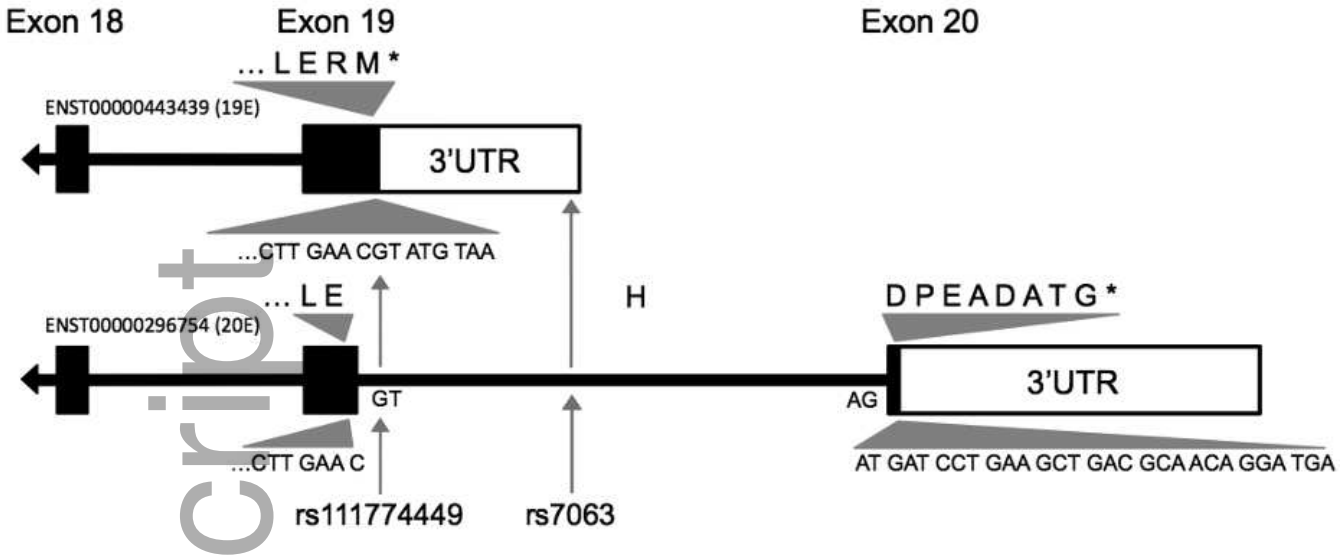
Fig.3: Genotypes across chromosome 5q15 locus SNPs correlate with expression of two isoforms of the *ERAP1* gene. (A) Transcript expression of *ERAP1* isoforms 19E (gray) and 20E (white) split by genotype at sQTL rs7063 (risk allele A). **(B)** *ERAP1* isoform 19E proportion split by genotype at rs7063. **(C)** Locus plot showing *P*-value for the effect of chromosome 5q15 locus SNPs on *ERAP1* isoform 19E proportion plotted against genomic region. Top sQTL rs7063 is marked in purple and SNPs are coloured according to LD with this variant as per the colour key at top right. Pie graphs show proportional contribution of each isoform to expression on the transcript level **(D)** and protein level **(E)** for individuals split by genotype at rs7063. The radius of each graph denotes mean total *ERAP1* expression on the transcript or protein level. **(F)** Protein expression of ERAP1 isoforms 19E (gray) and 20E (white) split by genotype at rs7063. **(G)** Total ERAP1 protein expression split by genotype at rs7063. n=73(AA), 43(AT) and 8(TT) for genotype groups in A and B, n=124 in the RNA-Seq cohort and n=39 in the mass spectrometry cohort in D and E, n=15(AA), 15(AT), 9(TT) for genotype groups in F and G.

Fig.4: Diagrammatic representation of the modeled structure of ERAP1 isoform 20E. The protein is shown as a cartoon colored blue, with the residues of the 20E extension shown

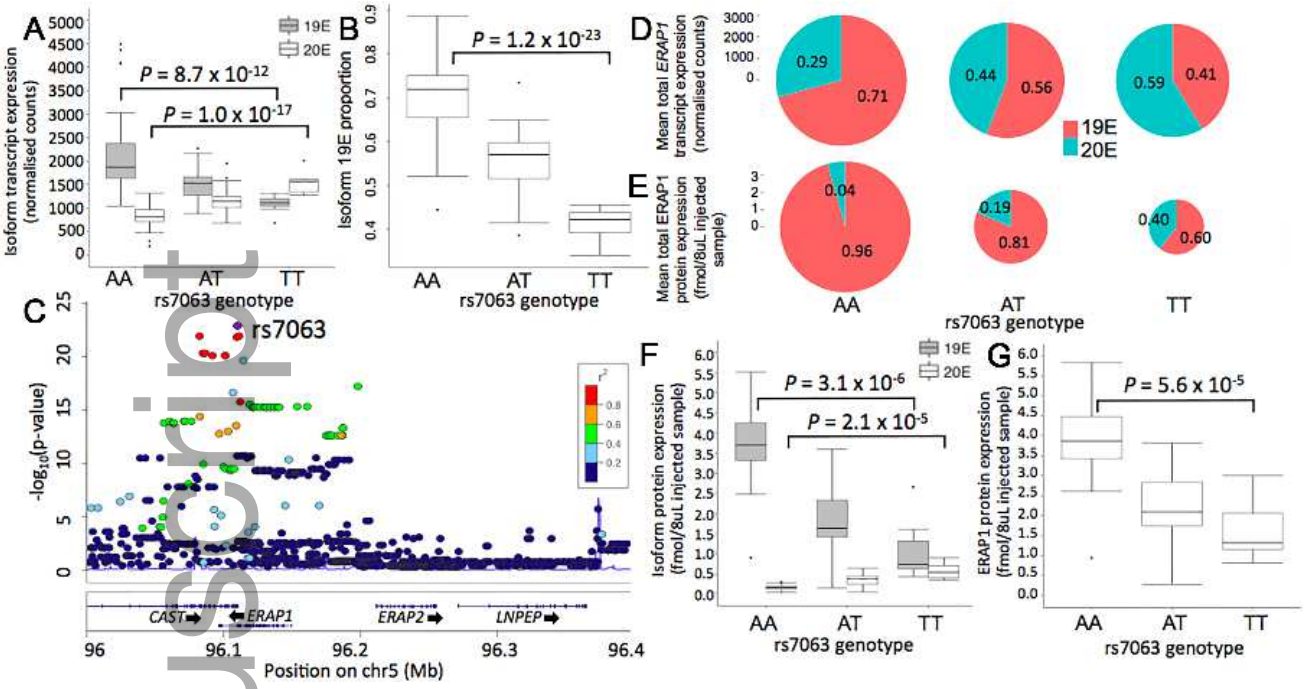
as bonds colored by atom type, with carbon atoms white. Residues in the body of the protein that are within contact with the modeled 20E extension are shown as bonds colored by atom, with carbon atoms green. The model suggests that the 20E extension residues sit across the surface of domain 4 of ERAP1 and are unlikely to interfere with protein function. (Inset) Overview of the model, with the protein shown as a cartoon colored in a rainbow gradient from the N-terminus (blue) to C-terminus (red). The region indicated by the circle is shown expanded in the main figure.



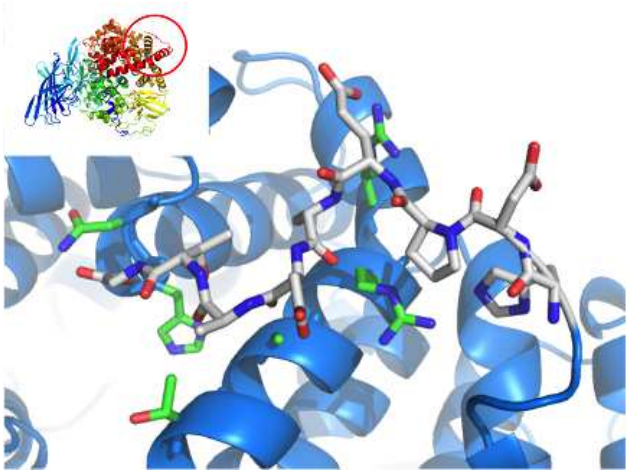
art_40369_f1.tif



art_40369_f2.tif



art_40369_f3.tif



art_40369_f4.tif

Author/s:
Hanson, AL; Cuddihy, T; Haynes, K; Loo, D; Morton, CJ; Oppermann, U; Leo, P; Thomas, GP; Kim-Anh, LC; Kenna, TJ; Brown, MA

Title:
Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile

Date:
2018-02-01

Citation:
Hanson, A. L., Cuddihy, T., Haynes, K., Loo, D., Morton, C. J., Oppermann, U., Leo, P., Thomas, G. P., Kim-Anh, L. C., Kenna, T. J. & Brown, M. A. (2018). Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile. ARTHRITIS & RHEUMATOLOGY, 70 (2), pp.255-265. https://doi.org/10.1002/art.40369.