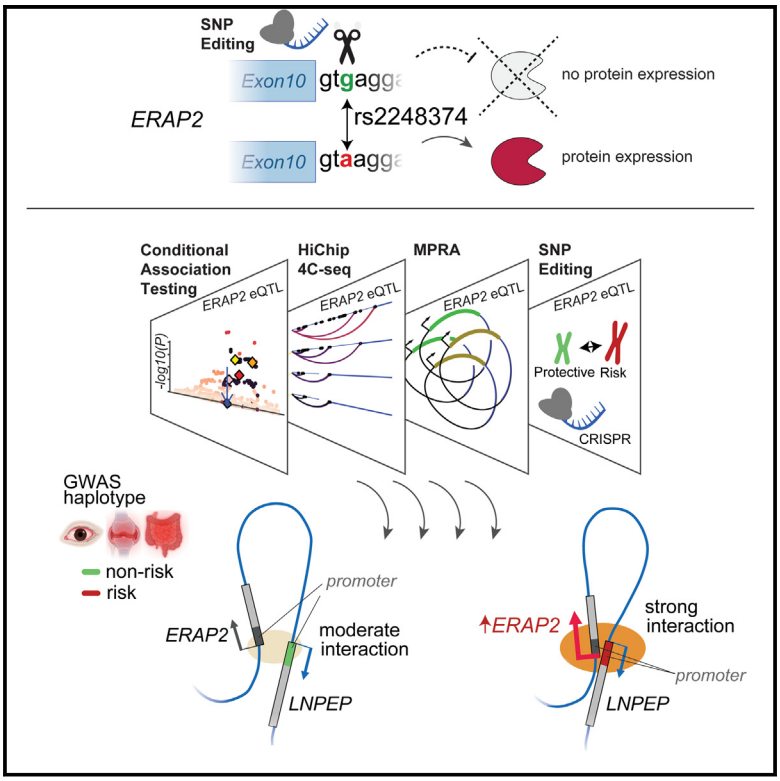
Cell Genomics

Article

A *cis*-regulatory element regulates *ERAP2* expression through autoimmune disease risk SNPs

Graphical abstract



Highlights

- *ERAP2* expression is critically dependent on the SNP rs2248374 near exon 10
- Autoimmune disease GWAS hits associate with ERAP2 levels independent of rs2248374
- Autoimmune risk SNPs downstream of *ERAP2* modify gene expression
- Autoimmune risk SNPs change local conformation and boost promoter interactions

Authors

Wouter J. Venema, Sanne Hiddingh, Jorg van Loosdregt, ..., Peter H.L. Krijger, Wouter de Laat, Jonas J.W. Kuiper

Correspondence

j.j.w.kuiper@umcutrecht.nl

In brief

ERAP2 gene variants are associated with autoimmune disorders and severe infectious diseases, but the function of these variants remains unknown. Venema et al. use genome editing and functional genomics to show that these genetic variants regulate *ERAP2* through multiple independent mechanisms, including by transforming a downstream gene promoter into an enhancer for *ERAP2*.

Venema et al., 2024, Cell Genomics *4*, 100460
January 10, 2024
https://doi.org/10.1016/j.xgen.2023.100460

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

A *cis*-regulatory element regulates *ERAP2* expression through autoimmune disease risk SNPs

Wouter J. Venema,[1,2] Sanne Hiddingh,[1,2] Jorg van Loosdregt,[2] John Bowes,[3] Brunilda Balliu,[4] Joke H. de Boer,[1] Jeannette Ossewaarde-van Norel,[1] Susan D. Thompson,[5] Carl D. Langefeld,[6] Aafke de Ligt,[1,2] Lars T. van der Veken,[7] Peter H.L. Krijger,[8] Wouter de Laat,[8] and Jonas J.W. Kuiper[1,2,9,*]

[1]Department of Ophthalmology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[2]Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[3]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK
[4]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
[5]Department of Pediatrics, University of Cincinnati College of Medicine, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[6]Department of Biostatistics and Data Science, and Center for Precision Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA
[7]Department of Genetics, Division Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[8]Oncode Institute, Hubrecht Institute-KNAW and University Medical Center Utrecht, 3584 CT Utrecht, the Netherlands
[9]Lead contact
*Correspondence: j.j.w.kuiper@umcutrecht.nl
https://doi.org/10.1016/j.xgen.2023.100460

SUMMARY

Single-nucleotide polymorphisms (SNPs) near the *ERAP2* gene are associated with various autoimmune conditions, as well as protection against lethal infections. Due to high linkage disequilibrium, numerous trait-associated SNPs are correlated with *ERAP2* expression; however, their functional mechanisms remain unidentified. We show by reciprocal allelic replacement that *ERAP2* expression is directly controlled by the splice region variant rs2248374. However, disease-associated variants in the downstream *LNPEP* gene promoter are independently associated with *ERAP2* expression. Allele-specific conformation capture assays revealed long-range chromatin contacts between the gene promoters of *LNPEP* and *ERAP2* and showed that interactions were stronger in patients carrying the alleles that increase susceptibility to autoimmune diseases. Replacing the SNPs in the *LNPEP* promoter by reference sequences lowered *ERAP2* expression. These findings show that multiple SNPs act in concert to regulate *ERAP2* expression and that disease-associated variants can convert a gene promoter region into a potent enhancer of a distal gene.

INTRODUCTION

MHC class I molecules (MHC-I) display peptides derived from intracellular proteins allowing $CD8^+$ T cells to detect infection and malignancy.[1,2] In the endoplasmic reticulum, aminopeptidases ERAP1 and ERAP2 shorten peptides that are presented by MHC-I.[3–5] Dysfunctional ERAP may alter the repertoires of peptides presented by MHC-I, potentially activating $CD8^+$ T cells and causing adverse immune responses.[6–8]

In genome-wide association studies (GWASs), polymorphisms at *5q15* (chromosome 5, *q* arm, G-band *15*) near the *ERAP1* and *ERAP2* genes have been associated with multiple autoimmune conditions. Among them are ankylosing spondylitis,[9,10] Crohn's disease (CD),[11] juvenile idiopathic arthritis (JIA),[12] birdshot chorioretinopathy (BCR),[13,14] psoriasis, and Bechet's disease.[15,16] The single-nucleotide polymorphisms (SNPs) identified in GWAS as disease risk SNPs in *ERAP1* usually correspond to changes in amino acid residues, resulting in proteins with different peptide trimming activities and expression levels.[8,17–20]

On the other hand, many SNPs near *ERAP2* are highly correlated with the level of *ERAP2* expression (i.e., expression quantitative trait loci [eQTLs] for *ERAP2*).[21,22] Due to linkage disequilibrium (LD) between these SNPs, there are two common *ERAP2* haplotypes; one haplotype encodes enzymatically active ERAP2 protein while the alternative haplotype encodes transcript with an extended exon 10 that contains premature termination codons, inhibiting mRNA and protein expression.[23] The haplotype that produces full-size *ERAP2* increases the risk of autoimmune diseases such as CD, JIA, and BCR, but it also protects against severe respiratory infections like pneumonia,[24] as well as historically the *Black Death*, caused by the bacterium *Yersinia*

pestis.[11–13,25] There is a SNP rs2248374 (allele frequency ~50%) located within a donor splicing site directly after exon 10 that tags these common haplotypes.[14,23,26] Consequently, rs2248374 is assumed to be the sole variant responsible for *ERAP2* expression. Although this is supported by association studies and minigene-based assays,[23,26] strikingly, there have been no studies evaluating *ERAP2* expression after changing the allele of this SNP in genomic DNA. This leaves the question of whether the rs2248374 genotype is essential for *ERAP2* expression unanswered.

More than a hundred additional *ERAP2* eQTLs located in and downstream of the *ERAP2* gene, form a large "extended ERAP2 haplotype."[13] It is commonly assumed that these *ERAP2* eQTLs work solely by tagging (i.e., in LD with rs2248374).[23,25,27–30] There is however, evidence that some SNPs in the extended *ERAP2* haplotype may influence *ERAP2* expression independent of rs2248374.[20,31] The use of CRISPR-Cas9 genome editing and functional genomics may be able to unravel the *ERAP2* haplotypes and identify causal variants that regulate *ERAP2* expression but are obscured by LD with rs2248374 in association studies.

We investigated whether rs2248374 is sufficient for the expression of *ERAP2*. Polymorphisms influencing *ERAP2* expression were identified using allelic replacement by CRISPR-mediated homologous repair and conformation capture assays. We report that rs2248374 was indeed critical for *ERAP2* expression but that *ERAP2* expression is further influenced by additional SNPs that facilitate a local conformation that increases promoter interactions.

RESULTS

ERAP2 expression depends on the genotype of rs2248374

The SNP rs2248374 is located downstream of exon 10 of *ERAP2* and its genotype strongly correlates with *ERAP2* expression. Predictions by deep neural network-based algorithms *SpliceAI* and *Pangolin* indicate that the A>G allelic substitution by rs2248374 inhibits constitutive splicing three base pairs (bp) upstream at the canonical exon-intron junction (*SpliceAI,* donor loss Δ score = 0.51, *Pangolin* Δ score = 0.58). Despite widespread assumption that this SNP controls *ERAP2* expression, functional studies are lacking.[23] Therefore, we first aimed to determine whether *ERAP2* expression is critically dependent on the genotype of this SNP. Allelic replacement by CRISPR-mediated homologous repair using a donor DNA template was used to specifically mutate rs2248374 G>A by homology directed repair (HDR) (Figure 1A; STAR Methods). Because HDR is inefficient,[32] a silent mutation was inserted into the donor template to produce a *TaqI* restriction site, which can be used to screen clones with correctly edited SNPs. As THP-1 cells are homozygous for the G allele of rs2248374 (Figure 1B), we used this cell line for experiments because it can be grown in single-cell-derived clones (see also Table S1). We targeted rs2248374 in THP-1 cells and established a clone that was homozygous for the A allele of rs2248374 (Figure 1B). Sequencing of the junctions confirmed that the integrations were seamless and precisely positioned in-frame.

SNP-array analysis was performed to exclude off-target genomic alterations giving rise to duplications and deletions in the genome of the gene edited cell lines (Figure S1; STAR Methods). We did not observe any of such unfavorable events. This confirmed that our editing strategy did not induce widespread genomic changes.[33] While THP-1 cells are characterized by genomic alterations, including large regions of copy number neutral loss of heterozygosity of chromosome 5 (including *5q15*)[34,35] (Figure S1), the results confirmed that single-cell clones from the unedited "wild-type" (WT, rs2248374-GG) THP-1 cells and "edited" THP-1 (rs2248374-AA) were genetically identical at *5q15*, which justifies their comparison (Figure 1C). In contrast with WT THP-1, *ERAP2* transcript became well detectable in THP-1 cells in which we introduced the A allele of rs2248374 (Figure 1D, see also Table S2). According to western blot analysis, WT THP-1 cells lack ERAP2 protein, while the rs2248374-AA clone expressed full-length ERAP2 (Figure 1E), which was enzymatically functional as determined by a fluorogenic *in vitro* activity assay (Figure 1F).

Oppositely, we then examined whether mutation of rs2248374 A>G would abolish *ERAP2* expression in cells naturally expressing ERAP2. The *Jurkat* T cell line was chosen because these cells are heterozygous for rs2248374 and naturally express ERAP2, and they possess the ability to grow in single-cell clones required to overcome the low efficiency of CRISPR knockin by HDR. To alter the single A allele of rs2248374 in the Jurkat cell line, we used a donor DNA template encoding the G variant (Figure S2A) and established a clone homozygous for the G allele of rs2248374 (Figure S2B). We found no changes between our unedited population and rs2248374 edited Jurkat cells at *5q15* by whole genome homozygosity mapping (Figure S2C). The A>G substitution at position rs2248374 depressed *ERAP2* mRNA expression (Figure S2D, see also Table S3) and abolished ERAP2 protein expression (Figure S2E). These results show that *ERAP2* mRNA and protein expression are critically dependent on the genotype of rs2248374 at steady-state conditions.

Disease risk SNPs are associated with *ERAP2* levels independent of rs2248374

Many additional SNPs at chromosome *5q15* show strong associations with *ERAP2* gene expression levels[36] (also known as *ERAP2* expression quantitative trait loci [eQTLs]). Despite LD between rs2248374 and the other *ERAP2* eQTLs, rs2248374 does not appear to be the strongest *ERAP2* eQTL in the GTEx database (data for GTEx "whole blood" are shown in Figure 2A, see also Table S4). Following this, we investigated the SNPs near the *ERAP2* gene that are associated with several T-cell-mediated autoimmune conditions, such as CD, JIA, and BCR (Tables S5–S7). We found strong evidence for colocalization between GWAS signals at *5q15* for BCR, CD, and JIA and *cis*-eQTLs for *ERAP2* (posterior probability of colocalization >90%) (Figures 2B–2D). This indicates that these SNPs alter the risk for autoimmunity through their effects on *ERAP2* gene expression. It is noteworthy, however, that the GWAS hits at *5q15* for CD, BCR, and JIA are in high LD ($r^2 > 0.9$) with each other but not in high LD with rs2248374 ($r^2 < 0.8$) (Figure 2E). Furthermore, the GWAS association signal at *5q15* for JIA that was obtained under a dominant model (lead variant rs27290; $P_{dominant} = 7.5 \times 10^{-9}$) did not include

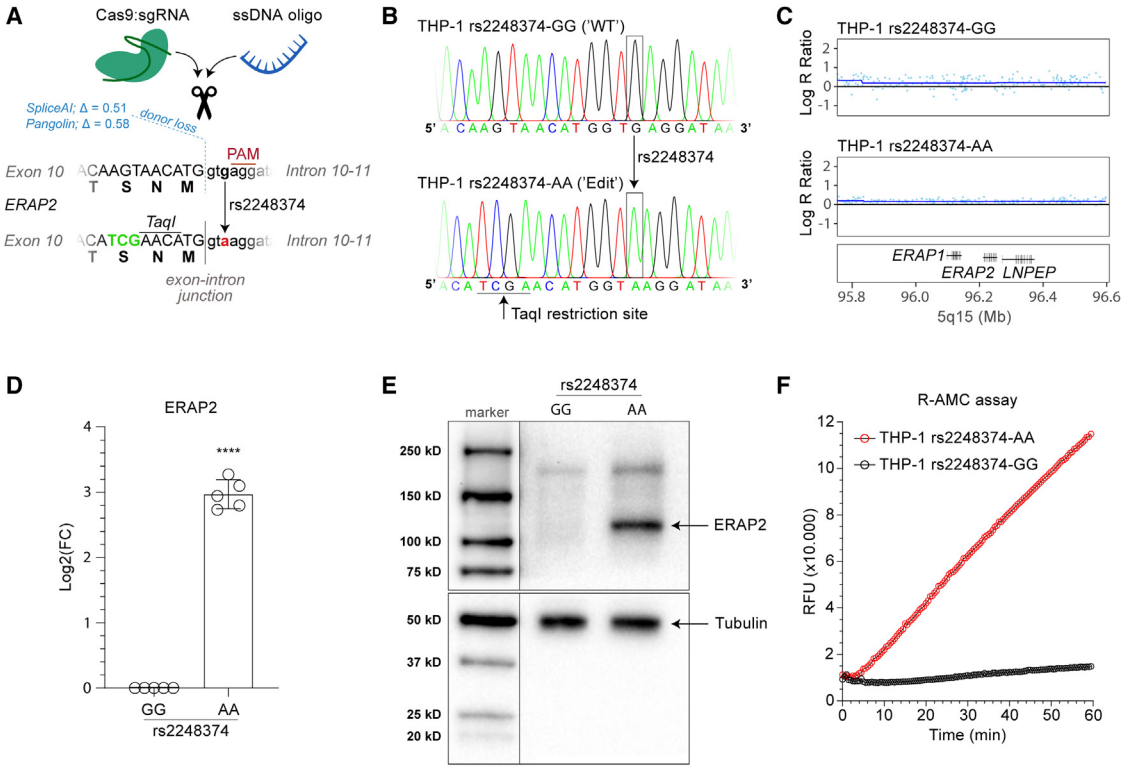


Figure 1. The A allele of rs2248374 is essential for full-length ERAP2 expression

(A) Overview of the CRISPR-Cas9-mediated homology directed repair (HDR) strategy for SNP allelic replacement of the G allele of rs2248374 to the A allele in THP-1 cells. The single-strand DNA oligo template introduces the A allele at position rs2248374, and a silent *TaqI* restriction site used for screening successfully edited clones. The predicted effect size (*delta* scores from *SpliceAI* and *Pangolin*, see STAR Methods) and intended position that exhibits altered splicing induced by the G allele of rs2248374 is shown in blue.

(B) Sanger sequencing data showing THP-1 "WT" with the single rs2248374-G variant and the successful SNP modification to the A allele of rs2248374.

(C) SNP-array-based copy number profiling and analysis of regions of homozygosity of unedited and edited THP-1 clones demonstrating no other genomic changes. Plot is zoomed in on *5q15*. Genome-wide results are outlined in Figure S1.

(D) *ERAP2* gene expression determined by qPCR in cellular RNA from five biological replicates of THP-1 cells unedited or edited for the genotype of rs2248374. The (****) indicates results from a t test, $p < 0.001$.

(E) Western blot analysis of ERAP2 protein in cell lysates from THP-1 cells unedited or edited for the genotype of rs2248374. Data show a single western blot analysis.

(F) Hydrolysis (expressed as relative fluorescence units [RFUs]) of the substrate L-Arginine-7-amido-4-methylcoumarin hydrochloride (R-AMC) by immunoprecipitated ERAP2 protein from THP-1 cell lines unedited or edited for the genotype of rs2248374. The generation of fluorescent AMC indicates ERAP2 enzymatic activity.

rs2248374 (JIA, $P_{dominant} = 0.65$), which indicates that the variants increase susceptibility to JIA by different mechanisms (Figure 2D). In line with this, we previously reported that the lead variant rs7705093 (Figure 2C) is associated with BCR after conditioning on rs2248374.[37] These findings reveal that SNPs implicated in these complex human diseases by GWAS may affect *ERAP2* expression through mechanisms other than rs2248374.

We therefore sought to determine if *ERAP2* eQTLs function independently of rs2248374. In agreement with the role ERAP2 plays in the MHC-I pathway that operates in most cell types, *ERAP2* eQTLs are shared across many tissues.[36,39] As a proof of principle, we used *ERAP2* eQTLs from RNA-sequencing data in whole blood from the GTEx Consortium[36] (Figure 2F). To test whether the disease-associated top association signals were independent from rs2248374, we performed conditional testing of the *ERAP2* eQTL signal by including the genotype of rs2248374 as a covariate in the regression model. Conditioning on rs2248374 revealed a complex independent *ERAP2* eQTL signal composed of many SNPs extending far downstream into the *LNPEP* gene. This secondary *ERAP2* eQTL signal included the lead variants at *5q15* for CD, BCR, and JIA ($P_{conditioned} < 4.8 \times 10^{-66}$), consistent with earlier findings[20,37] (Figure 2F, see also Table S8). We further strengthened these observations by using summary statistics from SNPs associated with plasma levels of ERAP2 from the INTERVAL study (called protein quantitative trait loci, or pQTLs).[38] After conditioning on rs2248374, among the top ERAP2 pQTLs in plasma was rs17486481 ($P_{conditioned} = 1.44 \times 10^{-275}$, see also Table S9), an intronic variant downstream of exon 12 that introduces a donor splice site leading to an uncharacterized alternatively spliced *ERAP2* transcript (termed "Haplotype C"),[4] but that is not in LD with any of the GWAS lead variants or with rs2248374 ($r^2 < 0.1$ in EUR), nor

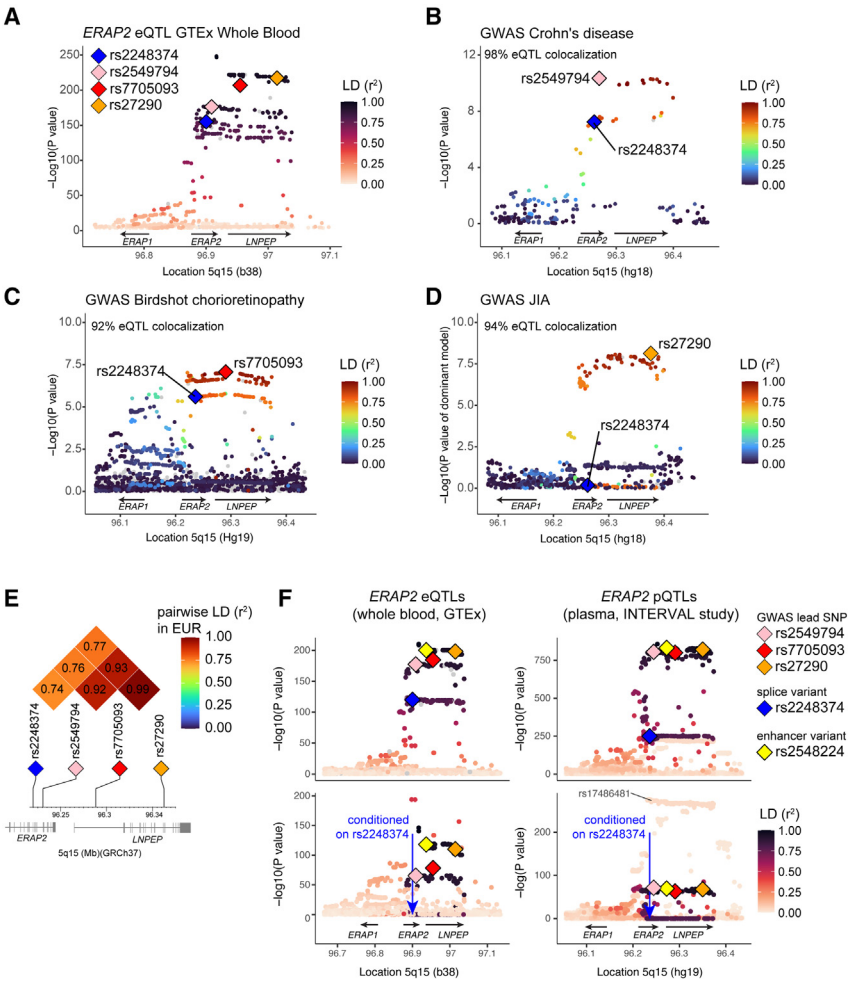


Figure 2. Autoimmune disease risk SNPs associated with ERAP2 levels independent from rs2248374 genotype

(A) *ERAP2* eQTL data from GTEx whole blood (Table S4). GWAS led variants at *5q15* for Crohn's disease (CD) (rs2549794, see B), birdshot chorioretinopathy (BCR) (rs7705093, see C), and juvenile idiopathic arthritis (JIA) (rs27290, see D) and rs2248374 are denoted by colored diamonds. The color intensity of each symbol reflects the extent of LD (r^2) from 1000 Genomes EUR samples with top *ERAP2* eQTL rs2927608. Gray dots indicate missing LD information.

(B–D) Regional association plots of GWAS from CD, BCR, and JIA (see also Tables S5–S7). For the CD we used the p value of rs2549782 (LD [r^2] = 1.0 with rs2248374 in EUR). The color intensity of each symbol reflects the extent of LD (r^2 estimated using 1000 Genomes EUR samples) with rs2927608. The results from colocalization analysis between GWAS signals and *ERAP2* eQTL data from whole blood (in A) is denoted.

(E) Pairwise LD (r^2 estimated using 1000 Genomes EUR samples) comparison between splice variant rs2248374 (*ERAP2*) and GWAS lead variants rs2549794 (CD), rs7705093 (BCR), and rs27290 (JIA).

(F) Initial association results and conditional testing of *ERAP2* eQTL data in whole blood from GTEx consortium (v8) and ERAP2 pQTL data from plasma proteomics of the INTERVAL study (see also Tables S8 and S9).[38] Conditioning on rs2248374 (dark blue diamond) revealed independent *ERAP2* eQTL and ERAP2 pQTL signals that include lead variants at *5q15* for CD, BCR, and JIA ($p < 5.0 \times 10^{-8}$). The human reference sequence genome assembly annotations are indicated.

associated with the here-studied autoimmune conditions. Regardless, in agreement with the mRNA data from GTEx, conditioning on rs2248374 revealed also strong independent association between GWAS lead variants and ERAP2 protein levels ($P_{conditioned} < 8.9 \times 10^{-64}$) (Figure 2F, see also Table S9). Based on these results, we conclude that GWAS signals at *5q15* are associated with ERAP2 levels independently of rs2248374.

SNPs in a downstream *cis*-regulatory element modulate *ERAP2* promoter interaction

Computational tools to predict the functional impact of non-coding variants may be highly inaccurate.[40] To prioritize likely causal variants by experimentally monitoring their effects on ERAP2, we aimed to resolve the function of SNPs that correlated with ERAP2 expression independent from rs2248374. First, we used CRISPR-Cas9 in *Jurkat* cells to eliminate a 116-kb genomic section containing most eQTLs downstream of *ERAP2* (which spans the entire *LNPEP* gene) (Figure S3A). We used Jurkat cells because these cells carry one chromosome with the protein-coding haplotype of *ERAP2* (Figure S2), so that we could screen for single-cell cultures that showed deletion of the region in the desired chromosome by genotyping the T allele of the *ERAP2* eQTL rs10044354 (LD [r^2] with rs7705093 in EUR = 0.98) located inside *LNPEP* by sanger sequencing. We identified a clone with evidence for deletion at *5q15*, and as confirmed by sanger sequencing (Figures S3B and S3C). A significant decrease in *LNPEP* mRNA levels by qPCR as well as depletion of the targeted region by whole genome zygosity mapping supported that we successfully depleted this region across chromosomes (Figures S3D and S3E). However, the *ERAP2* expression by qPCR was not significantly reduced by this approach (Figure S3E, see also Table S10). Close examination of the B allele frequency tracks of the SNP-array data revealed incomplete loss of heterozygosity for rs10044354 (and rs4360063, another *ERAP2* eQTL in full LD) indicating that we only achieved partial deletion of the region in the desired chromosome (Figure S4). Accordingly, we conclude that although we achieved modest depletion of the alternative alleles of eQTLs downstream of *ERAP2*, this was not sufficient to detect changes in mRNA levels.

Since allelic replacement would provide a more physiologically relevant approach, we next aimed to specifically alter the SNP alleles and evaluate the impact on *ERAP2* expression. The large size of the region containing all the "independent"

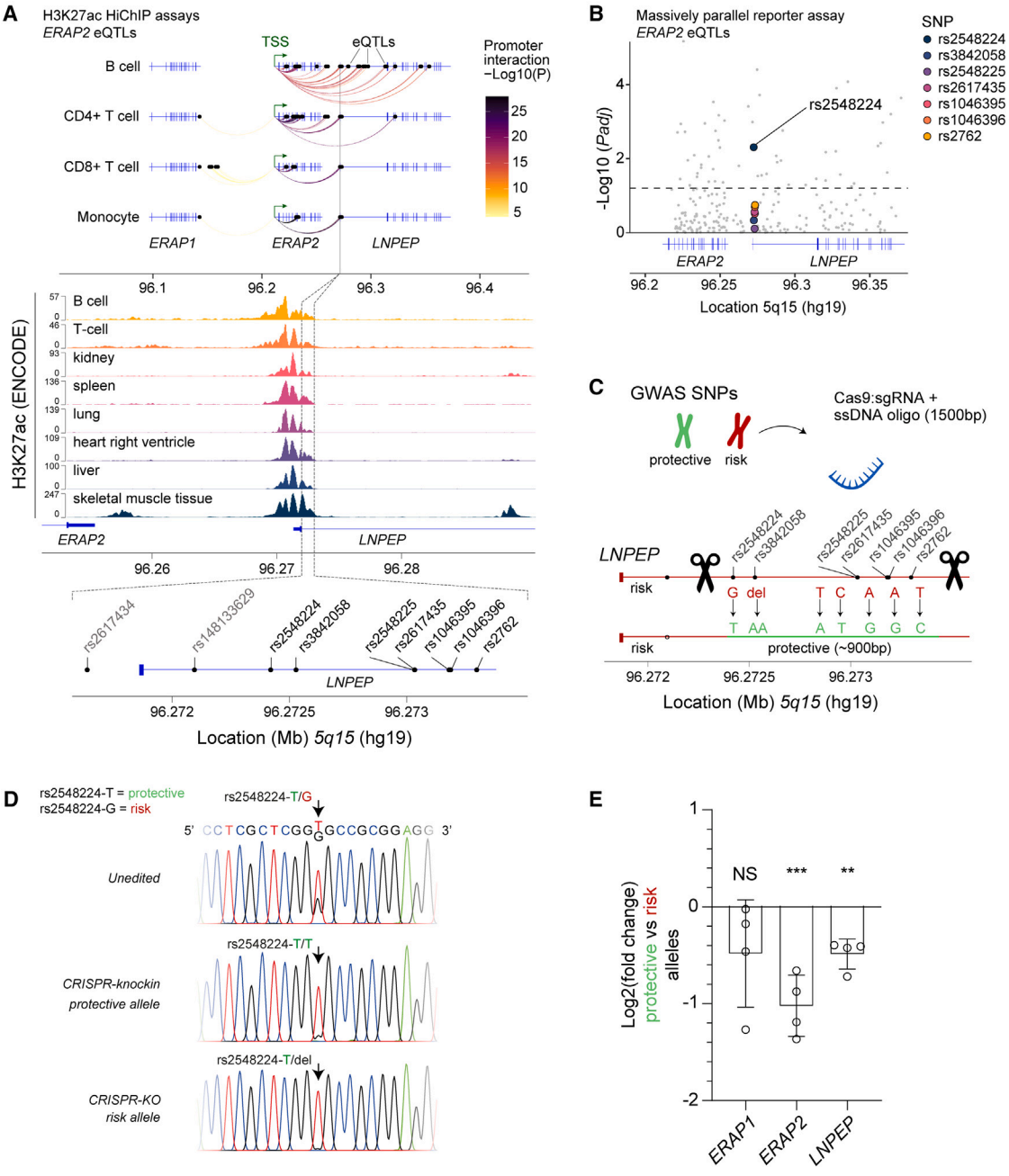


Figure 3. Autoimmune disease risk SNPs tag a downstream regulatory element that regulates *ERAP2* expression

(A) Chromosome conformation capture coupled with sequencing (Hi-C) data enriched by chromatin immunoprecipitation for the histone H3 lysine 27 acetylation (*H3K27ac*) in primary immune cells from Chandra et al.[42] Highlighted are the *ERAP2* eQTLs (black dots) that overlap with *H3K27ac* signals that significantly interact with the transcriptional start site of *ERAP2* in four different immune cell types (B cells, $CD4^+$ T cells, $CD8^+$ T cells, and monocytes). Nine common non-coding SNPs concentrated in an ~1.6-kb region exhibited strong interactions and overlay with *H3K27ac* signals from ENCODE data of heart, lung, liver, skeletal muscle, kidney, and spleen revealed.

(B) The −Log10(p values) (adjusted for multiple testing using the Benjamini-Hochberg method) of the effect of 986 *ERAP2* eQTLs on differential expressions (alternative versus reference allele) of their 150-bp window region from a massively parallel reporter assay as reported by Abell et al.[31] The seven SNPs identified by HiChIP in (A) are color-coded.

(C) Overview of the homology directed repair (HDR) strategy to use CRISPR-Cas9-mediated SNP replacement in Jurkat cells to switch the alleles from disease risk SNPs (i.e., alleles associated with higher *ERAP2* levels) to protective haplotype (i.e., alleles associated with lower *ERAP2* expression). The region from 5′ to 3′ spans 879 bp.

(legend continued on next page)

ERAP2 eQTLs prevents efficient HDR,[32,33] so we decided to prioritize a regulatory interval with *ERAP2* eQTLs. Genetic variation in non-coding enhancer sequences near genes can influence gene expression by interacting with the gene promoter.[41] Therefore, we leveraged chromosome conformation capture coupled with sequencing (Hi-C) data[42] enriched by chromatin immunoprecipitation for the activating histone H3 lysine 27 acetylation (*H3K27ac*, an epigenetic mark of active chromatin that marks enhancer regions) in primary T cells, B cells, and monocytes (STAR Methods), immune cells that share *ERAP2* eQTLs as shown by single-cell sequencing studies.[39] We selected *ERAP2* eQTLs located in active enhancer regions at *5q15* (i.e., H3K27ac peaks) that significantly interacted with the transcriptional start site of *ERAP2* for each immune cell type. This revealed diverse and cell-specific significant interactions of *ERAP2* eQTLs across the extended *ERAP2* haplotype in immune cells, indicating many regions harboring eQTLs that were physically in proximity with the transcription start site of *ERAP2* (Figure 3A). Note that none of these SNPs showed significant interaction with the promoters of *ERAP1* or *LNPEP*. Among these, nine common non-coding SNPs concentrated in an ~1.6-kb region downstream of *ERAP2* at the 5′ end of the gene body of *LNPEP* exhibited strong interactions with the *ERAP2* promoter (Figure 3A), suggesting that these SNPs lie within a potential regulatory element (i.e., enhancer) that is active in multiple cell lineages. Consistent with these data, examination of ENCODE data of heart, lung, liver, skeletal muscle, kidney, and spleen revealed enrichment of *H3K27ac* marks spanning the 1.6-kb locus, supporting that these SNPs lie within an *enhancer*-like DNA sequence that is active across tissues (Figure 3A). This also corroborates the finding that these SNPs are *ERAP2* eQTLs across tissues, as we showed previously[43] (Figure 2F). Data from a recent study[31] using targeted massively parallel reporter assays (MPRAs) support that this region may exhibit differential regulatory effects (i.e., altered transcriptional regulation), depending largely on the allele of SNP rs2548224 (difference in expression levels of target region; reference versus alternative allele for rs2548224, *Padj* = 4.9×10^{-3}) (Figure 3B). This SNP is also a very strong (rs2248374-independent) *ERAP2* eQTL and pQTL (Figure S5, see also Tables S8 and S9). In summary, this selected region downstream of *ERAP2* contained SNPs that are associated with ERAP2 expression independently of rs2248374, are physically in proximity with the *ERAP2* promoter (i.e., by Hi-C), and may exert allelic-dependent effects (i.e., by MPRA). Therefore, we hypothesized that the risk alleles of these SNPs associated with autoimmunity may increase the interaction with the promoters of *ERAP2*.

To investigate this, we first asked if specific introduction of the alternative alleles for these SNPs would affect the transcription of *ERAP2*. We targeted this region of the *ERAP2*-encoding chromosome in Jurkat cells using CRISPR-Cas9 and two guide RNAs in the presence of a large (~1,500 bp) single-stranded DNA template identical to the target region but encoding the alternative alleles for seven of the nine non-coding SNPs (Table 1). These SNPs were selected because they cluster close together (~900 bp distance from 5′ SNP rs2548224 to 3′ SNP rs2762) and are in tight LD (r^2~1 in EUR) with each other, as well as with the GWAS lead variants at *5q15* from CD, BCR, and JIA ($r^2 > 0.9$) (Figure S6). The introduction of the template DNA for CRISPR knockin by HDR did not induce other genomic changes (Figures 3C and S4). Sanger sequencing revealed targeting this intronic region by CRISPR-mediated HDR successfully altered the allele for SNPs rs2548224 in the regulatory element, but not the other targeted SNPs (Figure 3D, see also Figure S7). The single substitution of rs2548224 indicates that part of the repair template was used in the repair mechanisms, which is consistent with the observation that introduction of the substitution is generally highest at the positions close to the Cas9 cut site.[44] Regardless, altering the risk allele G to the reference allele T for rs2548224 resulted in significant decrease in *ERAP2* mRNA (unpaired t test, $p = 3.0 \times 10^{-4}$) (Figure 3E and Table S11). In agreement with the known ability of enhancers to regulate multiple genes within the same topologically associated domain, altering the alleles of these SNPs also resulted in significant reductions in the expression of the *LNPEP* gene (unpaired t test, p = 0.0018), but not *ERAP1* (Figure 3E). Last, to determine if the G allele of rs2548224 was sufficient by itself to induce *ERAP2* expression, we tested if altering the protective T allele to the risk G allele of rs2548224 affected *ERAP2* expression on a genetic background with otherwise protective alleles for all other *ERAP2* eQTLs (Figure S8A). To achieve this, we used our generated THP-1 rs2248374-AA clone (Figure 1) and successfully substituted the reference T allele to the disease risk allele G for rs2548224 using a 129-bp DNA repair template containing only this SNP (Figure S8B). The introduced risk G allele of rs2448224 did not result in a significant increase in the mRNA levels for *ERAP2* or *LNPEP* compared with clones with the reference T alleles (Figure S8C; Table S12). Overall, these results indicate that *ERAP2* gene expression can be downregulated by protective alleles of disease-associated SNPs downstream of the *ERAP2* gene in Jurkat cells, but not in THP-1 cells.

ERAP2 promoter contact is increased by autoimmune disease risk SNPs

RegulomeDB indicates that the SNP rs2548224 overlapped with 153 epigenetic mark peaks in various cell types (e.g., *POL2RA* in B cells). Considering its position within *LNPEP*'s promoter region, it makes it difficult to distinguish between local promoter and enhancer functions. To determine whether alleles of the SNPs in the regulatory element directly influenced contact with the *ERAP2* promoter, we used allele-specific 4C-seq in B cell lines generated from blood of three BCR patients carrying both the risk and non-risk allele (i.e., heterozygous for disease risk SNPs). Using nuclear proximity ligation, 4C-seq enables the quantification of

(D) Sanger sequencing results for the genotype of rs2548224 for Jurkat cells targeted by the CRISPR-based knockin approach outlined in (C). In comparison with unedited Jurkat cells and Jurkat cells in which the risk haplotype was deleted by CRISPR-Cas9-mediated knockout (as shown in Figure S3).
(E) Expression of *ERAP2*, *LNPEP,* and *ERAP1* by qPCR in Jurkat clones after allelic substitution of rs2548224. Data represent n = 4 biological replicates, Two-tailed unpaired t test was assessed to compare WT expression with the modified clone (**p < 0.01, ***p < 0.001).

Table 1. Details of the SNPs investigated in this study

SNP	Context in this study	Coord (GRChr37)	Alleles	MAF (EUR)	Distance from rs7705093	LD (D′)	LD (r^2)	Correlated alleles
rs2248374	*ERAP2* splice variant	chr5:96235896	(A/G)	0.4801	−54,751	0.99	0.75	C = G,T = A
rs2549794	lead SNP *5q15* in Crohn's disease[11]	chr5:96244549	(C/T)	0.4046	−46,098	0.98	0.92	C = T,T = C
rs2548224	*ERAP2* eQTL in regulatory region	chr5:96272420	(T/G)	0.4175	−18,227	0.99	0.98	C = T,T = G
rs3842058	*ERAP2* eQTL in regulatory region	chr5:96272528	(AA/−)	0.4175	−18,119	0.99	0.98	C = AA,T = −
rs2548225	*ERAP2* eQTL in regulatory region	chr5:96273033	(A/T)	0.4155	−17,614	1.0	0.98	C = A,T = T
rs2617435	*ERAP2* eQTL in regulatory region	chr5:96273034	(T/C)	0.4155	−17,613	1.0	0.98	C = T,T = C
rs1046395	*ERAP2* eQTL in regulatory region	chr5:96273180	(G/A)	0.4016	−17,467	1.0	0.93	C = G,T = A
rs1046396	*ERAP2* eQTL in regulatory region	chr5:96273187	(G/A)	0.4155	−17,460	0.99	0.98	C = G,T = A
rs2762	*ERAP2* eQTL in regulatory region	chr5:96273298	(C/T)	0.4145	−17,349	1.0	0.99	C=C,T = T
rs7705093	lead SNP *5q15* in birdshot chorioretinopathy[13]	chr5:96290647	(C/T)	0.4175	0	1.0	1.0	C=C,T = T
rs27290	lead SNP *5q15* in JIA[12]	chr5:96350088	(G/A)	0.4145	59,441	1.0	0.99	C = A,T = G

The minor allele frequency (MAF) and linkage disequilibrium (LD) for each SNP is indicated for the European (EUR) superpopulation of the 1000 Genomes.

contact frequencies between a genomic region of interest and the remainder of the genome.[45] Allele-specific 4C-seq has the advantage of measuring chromatin contacts of both alleles simultaneously and allows comparison of the risk allele versus the protective allele in the same cell population. We found that the downstream regulatory region formed specific contacts with the promoter of *ERAP2* (Figure 4A, see also Figure S9). Moreover, in two out of three patients, contact frequencies with the *ERAP2* promoter were substantially higher for the risk allele than the protective allele, supporting the idea that *ERAP2* expression may be a consequence of a direct regulatory interaction between the autoimmune risk SNPs and the gene promoter (Figure 4B, see also Figure S10).

DISCUSSION

In this study, we demonstrated that ERAP2 expression is initiated or abolished by the genotype of the common SNP rs2248374. Furthermore, we demonstrated that autoimmune disease risk SNPs identified by GWAS at *5q15* are statistically associated with *ERAP2* mRNA and protein expression independently of rs2248374. We show that autoimmune risk SNPs tag a gene-proximal DNA sequence that influences *ERAP2* expression and interacts with the gene's promoter more strongly if it encodes the risk alleles. Based on these findings, disease susceptibility SNPs at *5q15* likely do not confer disease susceptibility by alternative splicing, but by changing enhancer-promoter interactions of *ERAP2*.

The SNP rs2248374 is located at the 5′ end of the intron downstream of exon 10 of *ERAP2* within a donor splice region and strongly correlates with alternative splicing of precursor RNA.[23,26] While the A allele of rs2248374 results in constitutive splicing, the G allele is predicted to impair recognition of the motif by the spliceosome (Figure 1A), which is conceptually supported by reporter assays outside the context of the *ERAP2* gene.[26] Through reciprocal SNP editing in genomic DNA, we here demonstrated that the genotype of rs2248374 determines the production of full-length ERAP2 transcripts and protein.

Exon 10 is extended due to the loss of the splice donor site controlled by rs2248374 and consequently includes premature termination codons (PTCs) embedded in intron 10–11.[23,26] Transcripts that contain a PTC can in principle produce truncated proteins, but if translation terminates more than 50–55 nucleotides upstream ("50-55-nucleotide rule") of an exon-exon junction,[46] they are generally degraded through a process called *nonsense-mediated mRNA decay* (NMD). Our data show that *ERAP2* dramatically alters protein abundance proportionate to transcript levels, which is consistent with the notion that transcripts encoding the G allele of rs2248374 are subjected to NMD during steady state.[20,23] The loss of ERAP2 is relatively unusual, given that changes in ERAP2 isoform usage manifest so dramatically at the proteome level.[20,47] However, *ERAP2* transcripts can escape NMD under inflammatory conditions, such that haplotypes that harbor the G allele of rs2248374 have been shown to produce truncated ERAP2 protein isoforms,[29,48] not to be confused with "short" ERAP2 protein isoforms that are presumably generated by post-translational autocatalysis unrelated to rs2248374.[49]

Most protein-coding genes express one dominant isoform,[50] but since both alleles of rs2248374 are maintained at near equal

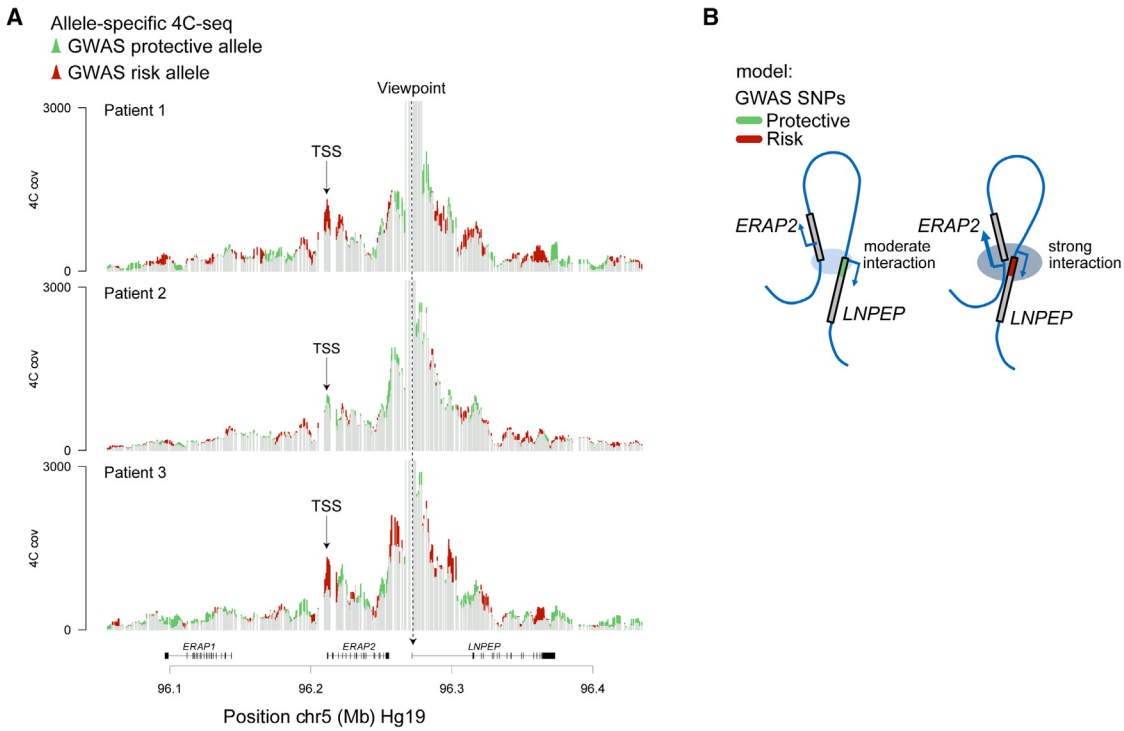


Figure 4. Autoimmune disease risk SNPs show high contact frequency with the *ERAP2* promoter in autoimmune patients

4C analysis of contacts between the downstream regulatory region across the *ERAP2* locus.

(A) 4C-seq contact profiles across the *ERAP2* locus in B cell lines from three patients with BCR that are heterozygous (e.g., rs2548224-G/T) for the *ERAP2* eQTLs located in the downstream regulatory element (the 4C viewpoint is centered on the SNP rs3842058 in the *LNPEP* promoter as depicted by the dashed line). The Y axis represents the normalized captured sequencing reads. The red lines in each track indicate the regions where the risk alleles show more interactions compared with the reference alleles, while the green lines indicate the regions where the reference alleles (i.e., protective alleles) show more interactions. TSS = transcription start site of *ERAP2*.

(B) Schematic representation of ERAP2 regulation by autoimmune risk SNPs in the downstream regulatory element showing the regulatory element with risk alleles (red) or reference (protective) alleles (green). The DNA region surrounding the *ERAP2* and *LNPEP* gene is shown in blue.

frequencies (allele frequency ~50%) in the human population, this leads to high interindividual variability in ERAP2 isoform profile.[23] ERAP2 may enhance immune fitness through balanced selection, especially since recent evidence indicates that the presumed "null allele" (i.e., the G allele of rs2248374) encodes distinct protein isoforms in response to infection.[29,51] A recent and unusual natural selection pattern during the *Black Death* for the haplotypes tagged by rs2248374 supports this,[25] as well as other studies of ancient DNA.[52,53] Nowadays, these haplotypes also provide differential protection against respiratory infections,[24] but they also modify the risk of modern autoimmune diseases like CD, BCR, and JIA. The SNP rs2248374 was long assumed to be primarily responsible for other disease-associated SNPs near *ERAP2*. Using conditional association analysis and mechanistic data, we challenged this assumption by showing that autoimmune disease risk SNPs identified by GWAS influence ERAP2 expression independently of rs2248374.

These findings are significant for two main reasons: First, these results demonstrate that chromosome structure plays important roles in the transcriptional control of *ERAP2* and thus that its expression is regulated by mechanisms beyond alternative splicing. We focused on a small *cis*-regulatory sequence downstream of *ERAP2* as a proof of principle. Here, we showed that disease risk SNPs alter physical interactions with the promoter in immortalized lymphoblast cell lines from autoimmune patients and that substitution of the allele of one common SNP (rs2548224) significantly affected the expression levels of *ERAP2*.

Another significant reason is that these findings have implications for our understanding of diseases in which ERAP2 is implicated. We recognize that the considerable LD between SNPs near *ERAP2* indicates that the effects of rs2248374 on splicing, as well as other mechanisms for regulation (i.e., chromosomal spatial organization), should often occur together. Because of their implications for the etiology of human diseases, it is still important to differentiate them functionally. Because disease-associated SNPs affect ERAP2 expression independently of rs2248374, ERAP2 may be implicated in autoimmunity not because it is expressed in susceptible individuals but because it is expressed at higher levels.[20,37] It corresponds with the notion that pro-inflammatory cytokines, such as interferons, upregulate ERAP2 significantly, while regulatory cytokines, like transforming growth factor β, downregulate it, or that ERAP2 is increased in lesions of autoimmune patients.[54,55] Overexpression of ERAP2 may be exploited therapeutically by lowering its

concentration in conjunction with local pharmacological inhibition of the enzymatic activity.[56]

Curiously, we note that the LD between rs2248374 and rs2548224 is higher in the African superpopulation of the 1000 Genomes compared with the European superpopulation (Figure S10), which is interesting considering the recent natural selection for these *ERAP2* variants in European populations.[52] Researchers have estimated that selection for rs2248374 and rs2548224 (proxy variant rs10044354 LD, r^2 = 0.99 in EUR) occurred in Europe within the past 2,000 years based on a large study of >2,000 ancient European genomes.[52,53] Of interest, the allele frequencies for these variants in contemporary African populations are very close to that of populations in Europe ~2,000 years ago (Figure S10).[4] Also, admixture events between archaic and modern European populations have introgressed variants in the *ERAP2* gene that are also predicted to affect expression and may influence ancestry-based structure of genetic variation in *ERAP2*.[57] To resolve evolutionary questions regarding selection for these variants further investigation is required that considers the full haplotypes of *ERAP2*.[4] For example, some amino acid variations in ERAP2 show substantial differences in frequency between European and other populations and are predicted to influence enzymatic function of ERAP2 that may modify the susceptibility to autoimmune diseases.[4,58]

Limitations of the study

We do like to stress that results from conditional eQTL and pQTL analysis in this study, supported by data from chromosome conformation capture coupled with sequencing analysis (Figure 4; [42]), as well as MPRA data[31] suggest that many more SNPs may act in concert to regulate *ERAP2* expression. A limitation of our work is that these SNPs have not all been independently examined. Also, there may be a cell-type-specific difference of *ERAP2* regulation since promoter-interacting eQTL data also indicate less significant interactions in monocytes than lymphocytes (Figure 3A). The observation that the haplotype tagged by rs2548224 (proxy variant rs2927608 in the study,[59] LD [r^2] = 0.95 in EUR) influences the transcriptional responses to influenza A virus in myeloid cells and not lymphocytes support potential cell-type-specific differences.[59] This is supported by the differences we noted in the rs2548224 allelic substitution between Jurkat (lymphocyte lineage) and THP-1 (myeloid lineage) cells. However, alternatively, it is also possible that the G allele is required in concert with other closely positioned *ERAP2* eQTLs that are in full LD to facilitate binding of transcription factors and increase expression levels and that substitution to T is sufficient to disrupt this process, but that the G allele is not sufficient to establish long-range chromatin contacts between the *LNPEP* promoter region and the *ERAP2* promoter by itself. Therefore, additional experimental work is needed to interrogate the extended *ERAP2* haplotype and follow up on some of the derived associations. Single-cell analysis shows that the many *ERAP2* eQTLs are shared between immune cells.[39,60] Mapping all the putative functional implications of these SNPs by CRISPR-based knockin experiments in genomic DNA is inefficient and labor-intensive, which makes their application in primary tissue challenging. MPRA provides a high-throughput solution to interrogating SNP effects, but lacks genomic context, and can only infer local allelic-dependent effects (i.e., no long-range interactions). Due to their dependency on PAM sequences for targeting regions of interest, CRISPR-Cas9-based enhancer-targeting systems[61] may not be able to dissect functional effects at a single nucleotide (i.e., SNP) resolution. It is possible to discern allelic-dependent effects in the canonical genomic context using allele-specific 4C sequencing, but in case of high LD and closely clustered SNPs (e.g., the ~900-bp region identified in this study) functional or non-functional SNPs cannot be distinguished within the sequence window of interest. Regardless, by integrating information from all these available technologies, we were able to shortlist an interval suitable for interrogation by CRISPR-based knockin techniques. A major drawback of this multi-step approach is that our study is therefore limited by sample size, and ideally, we should have successfully targeted the regulatory region in a larger number of cell lines. Also, while ERAP2 also shows tissue-shared genetic regulation, there may be important cell-type-specific regulatory mechanisms enforced by disease risk allele that require study of this mechanism in affected tissues and under inflammatory conditions. Finally, we have not functionally dissected all known haplotypes of ERAP2, such as haplotype C (tagged by splice variant rs17486481),[4] which was strongly associated with ERAP2 plasma levels after adjusting for rs2248374.

An enhancer-promoter loop increases transcriptional output through complex organization of chromatin, structural mediators, and transcription factors.[62–64] Although we narrowed down the *cis*-regulatory region to ~900 bp, the identity of the structural or transcriptional regulators that juxtapose this region with the *ERAP2* promoter remains elusive. Loop-forming transcription factors such as CTCF and protein analogues (e.g., YY1, the Mediator complex) have been shown to contribute to enhancer-promoter interactions.[64–67] Given that the here-identified *cis*-regulatory region is located within the *LNPEP* promoter, it is challenging to identify the factors responsible for *ERAP2* expression, since promoters are highly enriched for a large variety of transcription factor footprints (i.e., high chromatin immunoprecipitation sequencing [ChIP-seq] signals). Further studies are required to dissect how these *ERAP2* eQTLs modify enhancer activity and transcription, and how these mechanisms are distinguished from canonical promoter activity for *LNPEP* genes.

Conclusions

In conclusion, these results show that clustered genetic association signals that are associated with diverse autoimmune conditions and lethal infections act in concert to control expression of *ERAP2* and demonstrate that disease risk variants can convert a gene promoter region into a potent enhancer of a distal gene.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact

- Materials availability
- Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - CRISPR-Cas9-mediated allelic substitution
  - Allelic substitution experiments in THP-1 rs2248374-AA
  - *CRISPR-Cas9* knock-out of *ERAP2* eQTLs
  - Western Blot analysis
  - qPCR analysis
  - ERAP2 activity assay
  - Allele-specific 4C-seq
  - High-density SNP-array analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ERAP2 eQTL colocalization and conditional analysis
  - Splice prediction of rs2248374
  - ERAP2 promotor-interacting SNP selection
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100460.

ACKNOWLEDGMENTS

We thank Dr. Dennis C. Ko, Dr. Darragh Duffy, and Dr. Jimmie Ye for helpful discussions. J.K. is supported by UitZicht (project number 2018–1) and Stichting Lijf en Leven (project number 63).

AUTHOR CONTRIBUTIONS

J.K. initiated and coordinated the project. J.v.L. assisted with experimental design and supervised the project together with J.K. W.V., S.H., and A.L. performed SNP editing experiments. P.K. and W.d.L. performed and analyzed 4C-seq experiments. L.v.V conducted homozygosity experiments and analysis. J.B, S.D.T, and C.D.L performed SNP association testing in autoimmune diseases. B.B. performed eQTL analysis and assisted in the design of the computational methods used in the study. J.O.v.N. generated cell lines from patients and curated experimental data. W.V., J.d.B., and J.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used *WordTune* to improve readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Received: March 16, 2023
Revised: October 4, 2023
Accepted: November 9, 2023
Published: December 5, 2023

REFERENCES

1. Rock, K.L., Reits, E., and Neefjes, J. (2016). Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol. *37*, 724–737.
2. Cresswell, P., Ackerman, A.L., Giodini, A., Peaper, D.R., and Wearsch, P.A. (2005). Mechanisms of MHC class I-restricted antigen processing and cross-presentation. Immunol. Rev. *207*, 145–157.
3. Saveanu, L., Carroll, O., Lindo, V., Del Val, M., Lopez, D., Lepelletier, Y., Greer, F., Schomburg, L., Fruci, D., Niedermann, G., and van Endert, P.M. (2005). Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat. Immunol. *6*, 689–697.
4. Raja, A., and Kuiper, J.J.W. (2023). Evolutionary Immuno-Genetics of Endoplasmic Reticulum Aminopeptidase II (ERAP2). Genes Immun. https://doi.org/10.1038/s41435-023-00225-8.
5. Kuśnierczyk, P., and Stratikos, E. (2019). Endoplasmic reticulum aminopeptidases as a double-faced tool to increase or decrease efficiency of antigen presentation in health and disease. Hum. Immunol. *80*, 277–280.
6. López de Castro, J.A. (2018). How ERAP1 and ERAP2 Shape the Peptidomes of Disease-Associated MHC-I Proteins. Front. Immunol. *9*, 2463.
7. Cavers, A., Kugler, M.C., Ozguler, Y., Al-Obeidi, A.F., Hatemi, G., Ueberheide, B.M., Ucar, D., Manches, O., and Nowatzky, J. (2022). Behçet's disease risk-variant HLA-B51/ERAP1-Hap10 alters human CD8 T cell immunity. Ann. Rheum. Dis. *81*, 1603–1611.
8. Kuiper, J.J.W., and Venema, W.J. (2020). HLA-A29 and Birdshot Uveitis: Further Down the Rabbit Hole. Front. Immunol. *11*, 599558.
9. Robinson, P.C., Costello, M.E., Leo, P., Bradbury, L.A., Hollis, K., Cortes, A., Lee, S., Joo, K.B., Shim, S.C., Weisman, M., et al. (2015). ERAP2 is associated with ankylosing spondylitis in HLA-B27-positive and HLA-B27-negative patients. Annals of the rheumatic diseases *74*, 1627–1629.
10. Evans, D. M., Spencer, C. C., Pointon, J. J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A.,Pirinen, M., Stone, M. A., et alWellcome Trust Case Control Consortium 2 (WTCCC2) (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility.Nat. Genet., 43(8),761–767.
11. Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. *42*, 1118–1125.
12. Hinks, A., Cobb, J., Marion, M.C., Prahalad, S., Sudman, M., Bowes, J., Martin, P., Comeau, M.E., Sajuthi, S., Andrews, R., et al. (2013). Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat. Genet. *45*, 664–669.
13. Kuiper, J.J., Van Setten, J., Ripke, S., Van 'T Slot, R., Mulder, F., Missotten, T., Baarsma, G.S., Francioli, L.C., Pulit, S.L., De Kovel, C.G., et al. (2014). A genome-wide association study identifies a functional ERAP2 haplotype associated with birdshot chorioretinopathy. Hum. Mol. Genet. *23*, 6081–6087.
14. Gelfman, S., Monnet, D., Ligocki, A.J., Tabary, T., Moscati, A., Bai, X., Freudenberg, J., Cooper, B., Kosmicki, J.A., Wolf, S., et al. (2021). ERAP1, ERAP2, and Two Copies of HLA-Aw19 Alleles Increase the Risk for Birdshot Chorioretinopathy in HLA-A29 Carriers. Invest. Ophthalmol. Vis. Sci. *62*, 3.
15. Kirino, Y., Bertsias, G., Ishigatsubo, Y., Mizuki, N., Tugal-Tutkun, I., Seyahi, E., Ozyazgan, Y., Sacli, F.S., Erer, B., Inoko, H., et al. (2013). Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. Nat. Genet. *45*, 202–207.
16. Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2; Strange, A., Capon, F., Spencer, C.C.A., Knight, J., Weale, M.E., Allen, M.H., Barton, A., Band, G., Bellenguez, C., et al. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat. Genet. *42*, 985–990.

17. Hutchinson, J.P., Temponeras, I., Kuiper, J., Cortes, A., Korczynska, J., Kitchen, S., and Stratikos, E. (2021). Common allotypes of ER aminopeptidase 1 have substrate-dependent and highly variable enzymatic properties. J. Biol. Chem. *296*, 100443.

18. Evnouchidou, I., Kamal, R.P., Seregin, S.S., Goto, Y., Tsujimoto, M., Hattori, A., Voulgari, P.V., Drosos, A.A., Amalfitano, A., York, I.A., and Stratikos, E. (2011). Cutting Edge: Coding single nucleotide polymorphisms of endoplasmic reticulum aminopeptidase 1 can affect antigenic peptide generation in vitro by influencing basic enzymatic properties of the enzyme. J. Immunol. *186*, 1909–1913.

19. Tran, T.M., and Colbert, R.A. (2015). Endoplasmic reticulum aminopeptidase 1 and rheumatic disease: functional variation. Curr. Opin. Rheumatol. *27*, 357–363.

20. Hanson, A.L., Cuddihy, T., Haynes, K., Loo, D., Morton, C.J., Oppermann, U., Leo, P., Thomas, G.P., Lê Cao, K.A., Kenna, T.J., and Brown, M.A. (2018). Genetic Variants in ERAP1 and ERAP2 Associated With Immune-Mediated Diseases Influence Protein Expression and the Isoform Profile. Arthritis Rheumatol. *70*, 255–265.

21. de Castro, J.A.L., and Stratikos, E. (2019). Intracellular antigen processing by ERAP2: Molecular mechanism and roles in health and disease. Hum. Immunol. *80*, 310–317.

22. Zhou, Y.H., Gallins, P.J., Etheridge, A.S., Jima, D., Scholl, E., Wright, F.A., and Innocenti, F. (2022). A resource for integrated genomic analysis of the human liver. Sci. Rep. *12*, 15151.

23. Andrés, A.M., Dennis, M.Y., Kretzschmar, W.W., Cannons, J.L., Lee-Lin, S.Q., Hurle, B., NISC Comparative Sequencing Program; Schwartzberg, P.L., Williamson, S.H., Bustamante, C.D., et al. (2010). Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. PLoS Genet. *6*, e1001157.

24. Hamilton, F., Mentzer, A.J., Parks, T., Baillie, J.K., Smith, G.D., Ghazal, P., and Timpson, N.J. (2023). Variation in ERAP2 has opposing effects on severe respiratory infection and autoimmune disease. Am. J. Hum. Genet. *110*, 691–702.

25. Klunk, J., Vilgalys, T.P., Demeure, C.E., Cheng, X., Shiratori, M., Madej, J., Beau, R., Elli, D., Patino, M.I., Redfern, R., et al. (2022). Evolution of immune genes is associated with the Black Death. Nature *611*, 312–319.

26. Coulombe-Huntington, J., Lam, K.C.L., Dias, C., and Majewski, J. (2009). Fine-scale variation and genetic determinants of alternative splicing across individuals. PLoS Genet. *5*, e1000766.

27. Rotival, M., Quach, H., and Quintana-Murci, L. (2019). Defining the genetic and evolutionary architecture of alternative splicing in response to infection. Nat. Commun. *10*, 1671.

28. Yao, Y., Liu, N., Zhou, Z., and Shi, L. (2019). Influence of ERAP1 and ERAP2 gene polymorphisms on disease susceptibility in different populations. Hum. Immunol. *80*, 325–334.

29. Ye, C.J., Chen, J., Villani, A.C., Gate, R.E., Subramaniam, M., Bhangale, T., Lee, M.N., Raj, T., Raychowdhury, R., Li, W., et al. (2018). Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. Genome Res. *28*, 1812–1825.

30. Bossini-Castillo, L., Glinos, D.A., Kunowska, N., Golda, G., Lamikanra, A.A., Spitzer, M., Soskic, B., Cano-Gamez, E., Smyth, D.J., Cattermole, C., et al. (2022). Immune disease variants modulate gene expression in regulatory CD4+ T cells. Cell Genom. *2*, 100117.

31. Abell, N.S., DeGorter, M.K., Gloudemans, M.J., Greenwald, E., Smith, K.S., He, Z., and Montgomery, S.B. (2022). Multiple causal variants underlie genetic associations in humans. Science (New York, N.Y.) *375*, 1247–1254.

32. Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., and Ma, W. (2018). Methodologies for Improving HDR Efficiency. Front. Genet. *9*, 691.

33. Tian, R., Pan, Y., Etheridge, T.H.A., Deshmukh, H., Gulick, D., Gibson, G., Bao, G., and Lee, C.M. (2020). Pitfalls in Single Clone CRISPR-Cas9 Mutagenesis to Fine-map Regulatory Intervals. Genes *11*, 504.

34. Odero, M.D., Zeleznik-Le, N.J., Chinwalla, V., and Rowley, J.D. (2000). Cytogenetic and molecular analysis of the acute monocytic leukemia cell line THP-1 with an MLL-AF9 translocation. Genes, chromosomes & cancer *29*, 333–338.

35. Adati, N., Huang, M.C., Suzuki, T., Suzuki, H., and Kojima, T. (2009). High-resolution analysis of aberrant regions in autosomal chromosomes in human leukemia THP-1 cell line. BMC Res. Notes *2*, 153.

36. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

37. Kuiper, J.J.W., Setten, J.V., Devall, M., Cretu-Stancu, M., Hiddingh, S., Ophoff, R.A., Missotten, T.O.A.R., Velthoven, M.V., Den Hollander, A.I., Hoyng, C.B., et al. (2018). Functionally distinct ERAP1 and ERAP2 are a hallmark of HLA-A29-(Birdshot) Uveitis. Hum. Mol. Genet. *27*, 4333–4343.

38. Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., et al. (2018). Genomic atlas of the human plasma proteome. Nature *558*, 73–79.

39. Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M.G., Andersen, S., Lu, Q., Rowson, A., Taylor, T.R.P., Clarke, L., et al. (2022). Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science (New York, N.Y.) *376*, eabf3041.

40. Liu, L., Sanderford, M.D., Patel, R., Chandrashekar, P., Gibson, G., and Kumar, S. (2019). Biological relevance of computationally predicted pathogenicity of noncoding variants. Nat. Commun. *10*, 330.

41. Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310.

42. Chandra, V., Bhattacharyya, S., Schmiedel, B.J., Madrigal, A., Gonzalez-Colin, C., Fotsing, S., Crinklaw, A., Seumois, G., Mohammadi, P., Kronenberg, M., et al. (2021). Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. Nat. Genet. *53*, 110–119.

43. Lu, A., Thompson, M., Grace Gordon, M., Dahl, A., Ye, C.J., Zaitlen, N., et al. (2023). Fast and powerful statistical method for context-specific QTL mapping in multi-context genomic studies. Preprint at bioRxiv.

44. Boel, A., De Saffel, H., Steyaert, W., Callewaert, B., De Paepe, A., Coucke, P.J., and Willaert, A. (2018). CRISPR/Cas9-mediated homology-directed repair by ssODNs in zebrafish induces complex mutational patterns resulting from genomic integration of repair-template fragments. Dis. Model. Mech. *11*, dmm035352.

45. Krijger, P.H.L., Geeven, G., Bianchi, V., Hilvering, C.R.E., and de Laat, W. (2020). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. Methods (San Diego, Calif.) *170*, 17–32.

46. Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. *23*, 198–199.

47. Sulakhe, D., D'Souza, M., Wang, S., Balasubramanian, S., Athri, P., Xie, B., Canzar, S., Agam, G., Gilliam, T.C., and Maltsev, N. (2019). Exploring the functional impact of alternative splicing on human protein isoforms using available annotation sources. Briefings Bioinf. *20*, 1754–1768.

48. Tanioka, T., Hattori, A., Masuda, S., Nomura, Y., Nakayama, H., Mizutani, S., and Tsujimoto, M. (2003). Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases. J. Biol. Chem. *278*, 32275–32283.

49. Mattorre, B., Caristi, S., Donato, S., Volpe, E., Faiella, M., Paiardini, A., Sorrentino, R., and Paladini, F. (2022). A Short ERAP2 That Binds IRAP Is Expressed in Macrophages Independently of Gene Variation. Int. J. Mol. Sci. *23*, 4961.

50. Ezkurdia, I., Rodriguez, J.M., Carrillo-de Santa Pau, E., Vázquez, J., Valencia, A., and Tress, M.L. (2015). Most highly expressed protein-coding genes have a single dominant isoform. J. Proteome Res. *14*, 1880–1887.

51. Saulle, I., Vanetti, C., Goglia, S., Vicentini, C., Tombetti, E., Garziano, M., Clerici, M., and Biasin, M. (2020). A New ERAP2/Iso3 Isoform Expression Is Triggered by Different Microbial Stimuli in Human Cells. Could It Play a Role in the Modulation of SARS-CoV-2 Infection? Cells *9*, 1951.

52. Kerner, G., Neehus, A.L., Philippot, Q., Bohlen, J., Rinchai, D., Kerrouche, N., Puel, A., Zhang, S.Y., Boisson-Dupuis, S., Abel, L., et al. (2023). Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. Cell Genom. *3*, 100248.

53. Laval, G., Patin, E., Quintana-Murci, L., and Kerner, G. (2023). Deep estimation of the intensity and timing of selection from ancient genomes. Preprint at bioRxiv.

54. Berglund, A.K., Hinson, A.L., and Schnabel, L.V. (2023). TGF-β downregulates antigen processing and presentation genes and MHC I surface expression through a Smad3-dependent mechanism. Preprint at bioRxiv *6*. http://biorxiv.org/content/early/2023/02/01/2023.01.30.526196.abstract.

55. Marusina, A.I., Ji-Xu, A., Le, S.T., Toussi, A., Tsoi, L.C., Li, Q., Luxardi, G., Nava, J., Downing, L., Leal, A.R., et al. (2023). Cell-Specific and Variant-Linked Alterations in Expression of ERAP1, ERAP2, and LNPEP Aminopeptidases in Psoriasis. J. Invest. Dermatol. *143*, 1157–1167.e10.

56. Laura, M., Ronan, G., Vy, L.B., Valentin, G., Omar, C.A., Virgyl, C., Piveteau, C., Melissa, R., Charlotte, F., Sandrine, W., et al. (2021). Modulators of hERAP2 discovered by high-throughput screening. Eur. J. Med. Chem. *211*, 113053.

57. Brand, C.M., Colbran, L.L., and Capra, J.A. (2023). Resurrecting the alternative splicing landscape of archaic hominins using machine learning. Nat. Ecol. Evol. *7*, 939–953.

58. Johnson, M.P., Roten, L.T., Dyer, T.D., East, C.E., Forsmo, S., Blangero, J., Brennecke, S.P., Austgulen, R., and Moses, E.K. (2009). The ERAP2 gene is associated with preeclampsia in Australian and Norwegian populations. Hum. Genet. *126*, 655–666.

59. Aquino, Y., Bisiaux, A., Li, Z., O'Neill, M., Mendoza-Revilla, J., Merkling, S.H., Kerner, G., Hasan, M., Libri, V., Bondet, V., et al. (2023). Dissecting human population variation in single-cell responses to SARS-CoV-2. Nature *621*, 120–128.

60. Perez, R.K., Gordon, M.G., Subramaniam, M., Kim, M.C., Hartoularos, G.C., Targ, S., Sun, Y., Ogorodnikov, A., Bueno, R., Lu, A., et al. (2022). Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. Science (New York, N.Y.) *376*, eabf1970.

61. Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K.E., Ni, M., and Xu, J. (2020). Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. Nat. Commun. *11*, 485.

62. Hua, P., Badat, M., Hanssen, L.L.P., Hentges, L.D., Crump, N., Downes, D.J., Jeziorska, D.M., Oudelaar, A.M., Schwessinger, R., Taylor, S., et al. (2021). Defining genome architecture at base-pair resolution. Nature *595*, 125–129.

63. Nakahashi, H., Kieffer Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. *3*, 1678–1689.

64. Chakraborty, S., Kopitchinski, N., Zuo, Z., Eraso, A., Awasthi, P., Chari, R., Mitra, A., Tobias, I.C., Moorthy, S.D., Dale, R.K., et al. (2023). Enhancer-promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. Nat. Genet. *55*, 280–290.

65. Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

66. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

67. Luan, J., Vermunt, M.W., Syrett, C.M., Coté, A., Tome, J.M., Zhang, H., Huang, A., Luppino, J.M., Keller, C.A., Giardine, B.M., et al. (2022). CTCF blocks antisense transcription initiation at divergent promoters. Nat. Struct. Mol. Biol. *29*, 1136–1144.

68. Hill, J.T., Demarest, B.L., Bisgrove, B.W., Su, Y.C., Smith, M., and Yost, H.J. (2014). Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. Dev. Dynam. *243*, 1632–1636.

69. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.

70. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

71. Li, D., Hsu, S., Purushotham, D., Sears, R.L., and Wang, T. (2019). WashU Epigenome Browser update 2019. Nucleic Acids Res. *47*, W158–W165.

72. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535–548.e24.

73. 1000 Genomes Project Consortium; Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

74. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat. Biotechnol. *34*, 339–344.

75. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. Nucleic Acids Res. *50*, D988–D995.

76. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

77. Zeng, T., and Li, Y.I. (2022). Predicting RNA splicing from DNA sequence using Pangolin. Genome Biol. *23*, 103.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat polyclonal anti-LRAP/ERAP2 (1:2500)	R&D Systems	Cat# AF3830 RRID:AB_2099119
Sheep polyclonal anti-LNPEP (1:2500)	R&D Systems	Cat# AF6386 RRID:AB_10717574
Mouse monoclonal anti-α-Tubulin (clone DM1A) (1:5000)	Sigma-Aldrich	Cat# T9026 RRID:AB_477593
Chemicals, peptides, and recombinant proteins		
Alt-R® HDR Enhancer (to improve HDR (68))	Integrated DNA Technologies	Cat# 1081072
L-Arginine-7-amido-4-methylcoumarin hydrochloride	Sigma-Aldrich	Cat# A2027
Amersham ECL™ Prime Western Blotting	GE Healthcare	Cat# RPN2236
IGEPAL® CA-630 (NP40 substitute)	Sigma-Aldrich	Cat# I8896
cOmplete Protease Inhibitor Cocktail	Roche	Cat# 11873580001
Critical commercial assays		
AllPrep® DNA/RNA/miRNA Universal Kit	Qiagen	Cat# 80224
Infinium Human CytoSNP-850K v1.2 BeadChip	Illumina	Cat# 20025644
TaqMan® SNP Genotyping Assay for rs2548224	Thermo Fisher Scientific	Cat. #:435137
Invitrogen™ SuperScript™ IV Reverse Transcriptase	Thermo Fisher Scientific	Cat. #: 15317696
Deposited data		
Code used in the manuscript	This manuscript	https://doi.org/10.34894/AVHJ1F
qPCR data	This manuscript	https://doi.org/10.34894/AVHJ1F
R code for SNP-array plots	This manuscript	https://doi.org/10.34894/AVHJ1F
Experimental models: Cell lines		
Human: THP-1	ATCC	TIB-202™
Human: Jurkat	ATCC	TIB-152™
Human: Birdshot patient derived lymphoblastoid cells	This paper (see methods)	N/A
Oligonucleotides		
Alt-R® S.p. Cas9 Nuclease V3	Integrated DNA Technologies	CAT#1081059
Alt-R tracrRNA	Integrated DNA Technologies	CAT#1072532
Ultramer® DNA Oligo	Integrated DNA Technologies	Table S13
Megamer™ ssDNA Fragment	Integrated DNA Technologies	Table S14
Alt-R HDR Donor Oligo	Integrated DNA Technologies	Table S15
Alt-R crRNAs	Integrated DNA Technologies	Tables S13–S15
qPCR primers	Integrated DNA Technologies	Table S17
Software and algorithms		
R Studio	R.4.2.2.	https://cran.rstudio.com/
NxClinical software v6.0	Bionano genomics	N/A
BENCH Lab CNV	Agilent	N/A
R package *sangerseqR*	Hill et al.[68]	https://www.bioconductor.org/packages/release/bioc/html/sangerseqR.html
R package *coloc* (v.5.1.0.1)	Giambartolomei et al.[69]	https://cran.r-project.org/web/packages/coloc/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GCTA -COJO	Yang et al.[70]	https://yanglab.westlake.edu.cn/software/gcta/#Overview
Graphpad Prism	Graphpad Software Inc.	www.graphpad.com
pipe4C	Krijger et al.[45]	https://github.com/deLaatLab/pipe4C
H3K27 acetylation data from ENCODE	*WASHU Epigenome Browser.* Li et al.[71]	http://epigenomegateway.wustl.edu/browser/(direct link to used data at *5q15* via: https://tiny.one/2p92f7tb
SpliceAI	Jaganathan et al.[72]	https://spliceailookup.broadinstitute.org/
Deposited data		
Whole blood significant eQTLs *ERAP2* v8 used for the colocalization analysis.	GTEx project[36]	https://gtexportal.org/home/datasets
Whole blood eQTL analysis chr5 v8 used in approximate conditional analysis.	GTEx project[36]	https://storage.googleapis.com/gtex-resources/GTEx_Analysis_v8_QTLs/GTEx_Analysis_v8_EUR_eQTL_all_associations/Whole_Blood.v8.EUR.allpairs.chr5.parquet
Annotation of rsIDs. GTEx_Analysis_2017-06-05_v8_WholeGenomeSeq_838Indiv_Analysis_Freeze.lookup_table.txt.gz	GTEx project[36]	https://gtexportal.org/home/datasets
Plasma pQTL ERAP2 summary statistics	INTERVAL study[38]	https://app.box.com/s/u3flbp13zjydegrxjb2uepagp1vb6bj2
GWAS summary statistics Crohn's Disease	IBD-genetics[11]	https://www.ibdgenetics.org/uploads/cd-meta.txt.gz
Massive parallel sequence data results for *ERAP2* eQTLs	Abell et al.[31]	N/A
The promoter-interacting eQTL summary statistics for *ERAP2* determined by H3K27ac HiChIP assays in primary immune cells	Chandra et al.[42]	N/A
Genotype file [hg19] of 1000 Genomes Project used in ensemble data slicer[73] with region lookup parameter "5:96000000-98000000".	The 1000 Genomes Project[73]	http://hgdownload.cse.ucsc.edu/gbdb/hg19/1000Genomes/phase3/ALL.chr5.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz
Genotype file [b38] of 1000 Genomes Project used in ensemble data slicer[73] with region lookup parameter "5:96000000-98000000".	The 1000 Genomes Project[73]	http://hgdownload.cse.ucsc.edu/gbdb/hg38/1000Genomes/ALL.chr5.shapeit2_integrated_snvindels_v2a_27022019.GRCh38.phased.vcf.gz
Sample ID's from the EUR superpopulation of the 1000 Genomes, used for annotation of samples.	The 1000 Genomes Project[73]	http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/integrated_call_samples_v3.20130502.ALL.panel

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Dr. Jonas Kuiper (j.j.w.kuiper@umcutrecht.nl).

Materials availability

The CRISPR-Cas9 edited THP-1 and Jurkat cell lines generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

Analysis scripts, data and summary statistics of GWAS, eQTL and pQTL analyses are available via dataverseNL. doi: https://doi.org/10.34894/AVHJ1F,

Remaining data are available in the Document S1 and Tables S1, S2–S12 and S13–S17.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The THP1 cell line (ATCC, TIB-202 ECACC Cat# 88081201, RRID:CVCL_0006, monocyte isolated from peripheral blood from an acute monocytic leukemia patient) and Jurkat cell line (ATCC, Clone E6-1,TIB-152; established from the peripheral blood of a 14-year-old, male, acute T cell leukemia patient) were purchased from ATCC. Cell lines were cultured in Roswell Park Memorial Institute 1640 medium (RPMI 1640, Thermo Fisher Scientific) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Biowest Riverside) and 1% penicillin/streptomycin (Thermo Fisher Scientific). The authenticity of each cell line was monitored by genome-wide SNP-array analysis (for technical details see *High-density SNP-array analysis* below). On the SNP array, 41 SNPs from a 97-SNP fingerprints for cancer cell line authentication (COSMIC Cell Line repository, available via https://cancer.sanger.ac.uk) were present. Jurkat and THP-1 cell lines used in this study had genotypes identical to those in the COSMIC Cell Line repository (Table S1).

The cell lines were used for CRISPR/Cas9 editing and clones were confirmed to be free of major off-target editing by genome-wide copy number SNP-array analysis. Figure legends or results sections describe the sample size for each experiment. EBV-immortalised lymphoblastoid cell lines (LCL) were generated from peripheral blood mononuclear cells (PBMC) from 3 birdshot chorioretinopathy patients (#LCL_112062; white, female, 61 years old; #LCL_112061, white female, 64 years old; #11_12893, white female 59 years old). The integrity of the genomes of the LCLs was assessed with whole genome SNP-array analysis, showing no abnormalities (Figure S9). Patient-derived cell lines were cultured in RPMI 1640 supplemented with 10% FBS 1% penicillin/streptomycin (Thermo Fisher Scientific). This study was approved by the institutional review board of the University Medical Center Utrecht, The Netherlands.

METHOD DETAILS

CRISPR-Cas9-mediated allelic substitution

We used the Alt-R gRNA system (Alt-R CRISPR-Cas9 tracrRNA ligated to custom Alt-R CRISPR-Cas9 crRNA) together with recombinant Alt-R S.p. Cas9 Nuclease V3 (Integrated DNA Technologies) and custom Ultramer DNA Oligo (Integrated DNA Technologies) as donor templates to modify rs2248374 through homology directed repair (Table S13). The guide RNA with the recombinant Cas9 nuclease (RNP complex) was assembled by incubating the Alt-R tracrRNA with the custom crRNA (Tables 13–S15) at 95°C for 5 min (1:1 ratio), followed by cooling down at room temperature. The RNP complex was mixed with the Alt-R S.p. Cas9 Nuclease and Buffer R (Neon system), followed by 10 min incubation. The custom DNA template was added to the mixture after RNP assembly. THP-1 cells or Jurkat cells were electroporated with the Neon Transfection System (Thermo Fisher Scientific) (for Jurkat cells: protocol A in Table S16). After electroporation the cells were incubated overnight with antibiotic-free culture medium with 20 μM Alt-R HDR Enhancer (Integrated DNA Technologies). The next day, cells were diluted in 10x dilution steps until a final concentration of 30 cells/mL was reached (total volume 10 mL). Cells were seeded in multiple flat-bottom 96-well plates at 20 μL cell suspension per well (<1 cell/well) and transferred to 24 well plates once grown confluently. Confluent cultured clones were lysed (~80% of total volume) in RLT buffer (Qiagen, Cat# 1030963) and DNA was isolated using Qiagen AllPrep DNA/RNA/miRNA Universal Kit (Cat# 80224). The flanking sequence of rs2248374 was amplified by PCR (Forward primer: AGGGAAAGAGAAGAATTGGA; Reverse primer: TCTCTTTCCTGTAGTGATTC) and PCR products incubated with the TaqI-v2 (R0149S, New England Biolabs) restriction enzyme (15 min at 65°C). Next, the PCR products were loaded on a 1% agarose gel (agarose, Acros Organics; 10x TAE UltraPure, Invitrogen) to assess restriction by *TaqI*. Samples with a cleaved PCR product were prepared for sanger sequencing to validate in-frame integration of donor DNA and correct modification of rs2248374. CRISPR-Cas9-mediated haplotype substitution (i.e., disease risk alleles to protective alleles) was achieved by using a large single-strand DNA template[74] (Table S14, 1500 bp Megamer, Integrated DNA Technologies) with the alternative allele for SNPs (from '5 to 3') rs2548224, rs3842058, rs2548225, rs2617435, rs1046395, rs1046396, and rs2762 and two guide RNAs (assembled separately) targeting upstream of rs2548224 and downstream of rs2762 respectively (Table 1) to introduce two double-strand breaks. The two RNP complexes were transfected together with the Megamer in Jurkat cells (Table S16, Jurkat protocol B). Jurkat cells were incubated overnight with antibiotic-free culture medium with 20 μM Alt-R HDR Enhancer (Integrated DNA Technologies), followed by single-cell seeding. PCR (Forward primer: GTGGGCA GTGGGAAAGTTGG; Reverse primer: TGTCTCCAGCATCAACTCTGA) and restriction analysis by TaqI (TaqI-v2, cleaves PCR products containing the G allele of rs1046396 tagging the protective allele) was used to identify clones without cleaved PCR bands (i.e., loss of risk haplotype), and allelic substitutions were validated by sanger sequencing.[68]

Allelic substitution experiments in THP-1 rs2248374-AA

Allelic substitution of rs2548224 in THP-1 [rs2248374-AA] cells was conducted as described above, but with minor adaptations such as the use of custom Alt-R HDR Donor Oligo (Enhanced Ultramer DNA oligo, Table S15) (Integrated DNA Technologies). Allelic

substitution by CRISPR-Cas9 mediated HDR in THP-1 was done by using 2.0×10^5 cells for the electroporation with Cas9/guideRNA ribonucleoprotein. After DNA isolation of single-cell clones using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen), we used the TaqMan SNP Genotyping Assay for rs2548224 (ThermoFisher Scientific, Catalog number: 4351379) to screen for clones with allelic substitution for rs2458224. Clones that seemed to have successfully underwent allelic substitution (or 'positive' clones) were used for PCR (along with 3 negative clones as controls) (Forward primer: AGGAGCGTTAAGTTGTGGCA, Reverse primer: ACACTCAAC ACCGCAAAACG) and subsequent agarose gel electrophoresis and sanger sequencing for validation of the TaqMan assay results was done. After confirming the TaqMan Assay results by sanger sequencing, selected clones were counted and seeded at the same density (2.0×10^5 cells/ml, in a total volume of 20 mL in a T75 flask). After culturing the cells for three days, cells were harvested and lysed in RLT Buffer Plus (Qiagen, Inc. 1053393) and lysates stored at the −80 freezer until RNA isolation. RNA isolation was performed using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Subsequently, the isolated RNA was reverse transcribed to produce cDNA using Invitrogen SuperScript IV Reverse Transcriptase (Fisher Scientific, Catalog number: 15317696) and qPCR analysis was performed to quantify gene expression of *ERAP1*, *ERAP2*, and *LNPEP*.

CRISPR-Cas9 knock-out of *ERAP2* eQTLs

To delete a 116 kb region downstream of *ERAP2* in Jurkat cells, two RNP complexes were assembled in parallel using Alt-R gRNA system (Alt-R CRISPR-Cas9 tracrRNA ligated to custom Alt-R CRISPR-Cas9 crRNA) together with recombinant Alt-R S.p. Cas9 Nuclease V3 (Integrated DNA Technologies). RNPs with guide RNAs (guide RNA1; GGCATTCCTTAAGGGTATCA and guide RNA2; GCGTTGCTTCACATATAAGT, see for details Figure S3) were transfected by electroporation into Jurkat cells (Jurkat protocol B in Table S16). Following transfection, Jurkat cells were diluted to 16.67 cells/mL and seeded in 96 well flat bottom plates (<1 cell/well at 50 μL cell suspension/well) for single cell cultures. Jurkat clones were screened by PCR (Fw: GTCCTTTCGCTGCTGATTTG; Rv: AGGTCATTCCACCACTTCATTGT), The PCR products were loaded on a 1% agarose gel (agarose, Acros Organics; 10x TAE UltraPure, Invitrogen) with the anticipated PCR product size only detectable after deletion of the entire DNA region, which was verified by sanger sequencing, as well as SNP-array copy number profiling (see section *High-density SNP-array analysis* (Figures S3 and S4).

Western Blot analysis

Western blotting was used to determine the protein levels of ERAP2. NP40 lysis buffer was used to prepare total cell lysates (1% NP40, 135 mM NaCl, 5 mM EDTA, 20 mM Tris-HCl, pH = 7.4), complemented with cOmplete protease inhibitor cocktail (Roche). Protein lysates (20 μg/lane) were separated on a 4–20% Mini-PROTEAN TGX gel (Bio-Rad Laboratories) and transferred to a polyvinylidene difluoride membrane (Immobilon-P PVDF, Millipore). After blocking in 5% non-fat dry milk in TBST, membranes were probed overnight at 4°C with antibodies that recognize ERAP2 (AF3830, R&D Systems) or α-tubulin (T9026, Sigma-Aldrich). Following washing, membranes were incubated with either anti-goat or anti-mouse secondary antibodies conjugated to HRP (DAKO). Protein bands were detected with Amersham ECL Prime Western Blotting (RPN2236, GE Healthcare) on the ChemiDoc Gel Imaging System (Bio-Rad Laboratories).

qPCR analysis

Gene expression was quantified by using qPCR. For *ERAP2* we used primers (Table S17) which bind to exon 10 (Fw: CATTCGGATCCCAAGATGAC) and exon 11 (Rv: GGAGTGAACACCCGTCTTGT) to determine functional *ERAP2* transcript based on rs2248374 variant. Primers targeting the control gene *RPL32* were used to calculate relative expression. (Fw: AGGGTTCGTAGAAGATTCAAGG; Rv: GGAAACATTGTGAGCGATCTC). The nucleotide sequence for the primers used for qPCR of *ERAP1* and *LNPEP* are shown in Table S17.

ERAP2 activity assay

To assess ERAP2 enzymatic function in THP-1 cells after introduction of the A allele of rs2248374, total cellular ERAP2 protein was enriched by immunoprecipitation and incubated with L-Arginine-7-amido-4-methylcoumarin hydrochloride (R-AMC, A2027, Sigma-Aldrich). Briefly, THP-1 cells were lysed (50 mM Tris, 150 mM NaCl, 1% Triton X-100, pH 7.5) and 500 μg cell lysate was incubated for 2h at 4°C with Protein-G-sepharose beads (BV-6511, BioVision) and anti-ERAP2 (AF3830, R&D Systems). Beads were washed in assay buffer (50 mM Tris, 1 mM DTT, pH 7.5) and resuspended in 100 μL assay buffer. Duplicate samples containing 40 μL beads solution were incubated with 40 μM of R-AMC (total volume 50 μL) for 1h at 37°C, while the fluorescent signal was measured (E_x: 370-10, E_m: 440-20) with a CLARIOstar plate reader (BMG Labtech).

Allele-specific 4C-seq

4C template preparation was performed as described.[45] Birdshot uveitis patient derived LCLs were cultured in RPMI 1640 supplemented with 10% FBS 1% penicillin/streptomycin (Thermo Fisher Scientific), washed and cross-linked in 2% formaldehyde. DNA was digested *in situ* with MboI (NEB) and Csp6I (Thermo Scientific). Primers used for inverse PCR (5′-TACACGACGCTCTTCC GATCTTGTGGTTGGTTACTTCAGGT-3′, 5′-ACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACCCTTGACATTAACAAG-3′) allow reading of the *ERAP2* eQTL rs3842058 (indel within the regulatory element which is located near rs2548224) in the viewpoint fragment that is associated with the risk allele. Products were sequenced using Illumina sequencing (Illumina NextSeq 500). 4C-seq reads were demultiplexed by matching the 5′-ends of the R1 reads to the reading primer sequence (allowing 2 mismatches) and split into risk and

non-risk reads based on the presence of the indel and its flanking sequence (GGACCCTTGACATTAACAAGTCTGAC for the risk allele and GGACCCTTGACATTAACAAGTCTTTG for the non-risk allele). Finally reads were mapped to the hg19 reference genome and processed using pipe4C[45] (github.com/deLaatLab/pipe4C) with the following parameters: normalization to 1 million reads in cis, window size 21, removal of top 2 read counts. Overlay plots were generated using R.

High-density SNP-array analysis

Genomic DNA of unedited and edited clones of THP-1 and Jurkat cell lines, and patient-derived LCLs was used for SNP-array copy number profiling and analysis of regions of homozygosity with the Infinium Human CytoSNP-850K v1.2 BeadChip (Illumina, San Diego, CA, USA). This array has ~850,000 single nucleotide polymorphisms (SNPs) markers across the genome and can detect genomic insertions and deletions as well as stretches of homozygosity. Data analysis was conducted using NxClinical software v6.0 (Bionano genomics, San Diego, CA, USA). Human genome build Feb. 2009 GRCh37/hg19 was used. Results were classified with BENCH Lab CNV software (Agilent, Santa Clara, CA, USA). The genotype data were used for cell line verification and monitoring for potential genomic alterations in single cell cultures.

QUANTIFICATION AND STATISTICAL ANALYSIS

ERAP2 eQTL colocalization and conditional analysis

Summary statistics of GWAS (see also *Data Availability*) from CD,[11] JIA,[12] and BCR[13] were used in visualisation of disease associated SNPs at *5q15*. Significant single tissue eQTLs for *ERAP2* (ENSG00000164308.16) in 'whole blood' were downloaded from the GTEx (v8) portal[36] and used for GWAS colocalization analysis with the *coloc* R package (STAR Methods).[69] We assessed the likelihood of colocalization using the posterior probability for hypothesis 4 ([PP4 =] that there is an association between both traits, and they are driven by the same causal variant(s)). The likelihood of colocalization was deemed high for associations with PP4 > 0.8. The summary statistics from the BCR GWAS was filtered for variants at (Hg19) chromosome 5 (chr5) from position 96000000–97200000 (n = 6307, Table S5). For colocalization analysis, we cross referenced these BCR variants with GTEx *ERAP2* eQTLs to obtain a final set of 679 SNPs. The CD GWAS summary statistics[11] (STAR Methods) were filtered for variants at (Hg18) chr5 from position 96081460–96461460 (n = 197, Table S6) of which 195 overlapped with GTEx *ERAP2* eQTLs for colocalization analyses. Because rs2248374 was not in the summary statistics of CD, we used the association statistics for variant rs2549782 (LD (r^2) = 1.0 in EUR of 1000 Genomes) for this variant. We filtered the GWAS summary statistics of JIA as reported by Hinks et al.[12] for variants at (Hg18) chr5 from position 95159342–98830661 (n = 974, Table S7) of which 495 variants were in common with the GTEx *ERAP2* eQTL data from whole blood and used in colocalization analysis. Note that the GWAS summary statistics reported by Hinks et al.[12] were calculated under a dominant genetic model. To identify *ERAP2* eQTLs and pQTLs with association signals independent from rs2248374, we used approximate conditional analysis by GCTA-COJO[70] using GCTA v1.94.1 and the '–cojo-cond' function (STAR Methods). For analysis of *ERAP2* eQTLs (Table S8), we used linkage disequilibrium (LD) information from the EUR superpopulation of the 1000 Genomes Project[72]. The genotype data for the EUR superpopulation was extracted using the ensemble data slicer[75] (STAR Methods). Sample IDs from the EUR superpopulation of the 1000 Genomes were obtained by filtering using meta-data from the 1000 Genomes ftp sever (STAR Methods) and converted into a.vcf files using the *vcfR R* package and formatted in Plink 1.09[76] by the ''–recode'' option into a BED file. We calculated the allele frequency of each SNP using the –freq function in Plink and LD with rs2927608 (i.e., primary GTEx *ERAP2* eQTL in whole blood) using the ''–r2'' option. The summary statistics for *ERAP2* pQTL data from blood plasma of the INTERVAL study as reported by Sun et al.[38] was used for conditional analysis of ERAP2 pQTLs (Table S9, and STAR Methods).

Splice prediction of rs2248374

We used the deep neural network *SpliceA*[72] and *Pangolin*[77] to predict the effects of the A>G allele substitution in pre-mRNA transcript sequences. Masked scores were predicted via https://spliceailookup.broadinstitute.org/ using "chr5 96900192 A G" as input for hg38, and "chr5 96235896 G A" for hg19.

ERAP2 promotor-interacting SNP selection

The promoter-interacting eQTL summary statistics for *ERAP2* determined by H3K27ac HiChIP assays in primary immune cells were obtained from Chandra et al.[42] We selected *ERAP2* eQTLs located in active enhancer regions at *5q15* (i.e., H3K27ac peaks) that significantly interacted with the transcriptional start site of *ERAP2* for each immune cell type ($p < 0.05$). We visualised *H3K27* acetylation data from primary tissue analysis from ENCODE using the *WASHU Epigenome Browser*[71] (v54.00) at http://epigenomegateway.wustl.edu/browser/(direct link to used data at *5q15* via: https://tiny.one/2p92f7tb). Massive parallel sequence data results for *ERAP2* eQTLs were obtained from supplementary data from Abell et al.[31]

ADDITIONAL RESOURCES

There are no additional resources.