

REVIEW ARTICLE OPEN



Evolutionary immuno-genetics of endoplasmic reticulum aminopeptidase II (ERAP2)

Aroosha Raja¹ and Jonas J. W. Kuiper¹

© The Author(s) 2023

Endoplasmic reticulum aminopeptidase 2 (ERAP2) is a proteolytic enzyme involved in adaptive immunity. The *ERAP2* gene is highly polymorphic and encodes haplotypes that confer resistance against lethal infectious diseases, but also increase the risk for autoimmune disorders. Identifying how ERAP2 influences susceptibility to these traits requires an understanding of the selective pressures that shaped and maintained allelic variation throughout human evolution. Our review discusses the genetic regulation of haplotypes and diversity in naturally occurring ERAP2 allotypes in the global population. We outline how these *ERAP2* haplotypes evolved during human history and highlight the presence of Neanderthal DNA sequences in *ERAP2* of modern humans. Recent evidence suggests that human adaptation during the last ~10,000 years and historic pandemics left a significant mark on the *ERAP2* gene that determines susceptibility to infectious and inflammatory diseases today.

Genes & Immunity (2023) 24:295–302; <https://doi.org/10.1038/s41435-023-00225-8>

INTRODUCTION

Endoplasmic reticulum aminopeptidase 2 (ERAP2) is a aminopeptidase located in the endoplasmic reticulum lumen that primarily functions in antigen processing for presentation by the class-I major histocompatibility complex (MHC-I) to initiate immune responses to infected cells [1]. Polymorphic variation in the form of single nucleotide variants (SNVs) in the *ERAP2* gene has been associated with increased susceptibility to chronic inflammatory disorders, such as *Crohn's disease*, *birdshot chorioretinopathy*, and *ankylosing spondylitis*, as well as with protection against severe infections such as, pneumonia, as well as the bubonic plague (i.e., *The Black Death*) in history [2–5]. SNVs located in and downstream of the *ERAP2* gene form extended *ERAP2* haplotypes [6] that display dramatic differences in transcriptional expression, but there remains little understanding how these functionally distinct haplotypes arose and mechanistically modify disease susceptibility in the human population. While the *ERAP2* gene exhibits signatures of *long-standing balancing selection* [6], new studies indicate also recent genetic adaptation in Europeans. In this review, we summarise the evolutionary history of human *ERAP2* haplotypes and discuss how natural selection helped shape allelic diversity that influences susceptibility to chronic inflammatory disease and infection resistance.

BIOLOGICAL FUNCTIONS OF ERAP2

The biological functions of ERAP2 in health and disease have been reviewed elsewhere [7–9] and will be covered only briefly here. The *ERAP2* gene (43 kb in length) clusters together with the *ERAP1* (47 kb) and *LNPEP* (102 kb) on chromosome *5q15* which based on close sequence homology constitute a functional subfamily of mammalian zinc-containing aminopeptidases (or

M1 metallopeptidases) that cleave N-terminal amino acids of bioactive and antigenic peptides [10, 11]. The evolutionary origins of these aminopeptidases remains uncertain, but others have speculated that *ERAP2* resulted from gene duplication from adjacent genes, such as *ERAP1* [6, 12]. As a M1 metallopeptidase, the 960 amino acid-long ERAP2 protein contains a GAMEN motif at amino acid position 334 (interacts with the N-terminus of peptide substrates) and a zinc-binding site encoded by amino acid motif HELAH (HEXXH-motif) at amino acid residues 370 (encoded by exon 6) [13, 14]. In vitro digestion analysis of purified ERAP2 has shown the capacity of the proteolytic enzyme to hydrolyze amino acid residues from the N-terminus of naturally occurring angiotensinergic and vascular inflammatory peptides, such as kallidin and angiotensin III (Fig. 1) [14]. Genome-wide association studies linking genetic variants in *ERAP2* to blood pressure [15] further support that ERAP2 may play a role in the renin-angiotensin system [16], however these functions remain largely speculative and require more investigation. Research on ERAP2 has been primarily focused on its role in antigen presentation [8, 9]. In the luminal site of the endoplasmic reticulum, ERAP2 is involved in the processing of N-terminal extended precursor peptides which determines their binding to MHC-I for presentation to CD8 + T cells at the cell surface (Fig. 1) [1, 14]. ERAP2 has been shown to generate the mature epitopes of immunogenic epitopes and directly enhance T cell recognition [17]. Mechanistically, ERAP2 shapes the immunopeptidome of the cell by the selective removal and indirect promotion of submotifs of peptides presented by MHC-I [18]. Due to its genetic association with several autoimmune diseases, ERAP2 may promote T-cell recognition of autoantigens, which contributes to destructive inflammation [7, 19, 20].

¹Department of Ophthalmology, Center for Translational Immunology, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands.

email: jj.w.kuiper@umcutrecht.nl

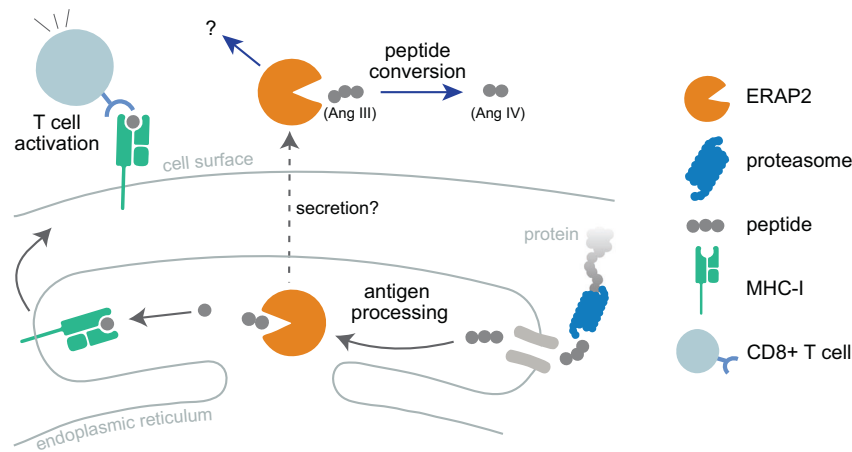


Fig. 1 The biological functions of ERAP2. Protein fragments (peptides) generated by cellular proteasomes are transported into the endoplasmic reticulum. ERAP2 is an aminopeptidase that trims N-terminal amino acid residues from peptides in the endoplasmic reticulum to prevent or facilitate binding to the major histocompatibility complex class I (MHC-I). The MHC-I peptide complexes are then presented on the surface of the cell to CD8 + T cells, allowing the immune system to identify and respond to infected or abnormal cells. The ERAP2 enzyme has also been shown to remove amino acid residues from naturally occurring signalling peptides, including the conversion of angiotensin derivatives [Angiotensin II/III into IV], indicating a potential role for ERAP2 in renin-angiotensin signalling. Additional unidentified pathways may be regulated by extracellular or secreted ERAP2 proteins.

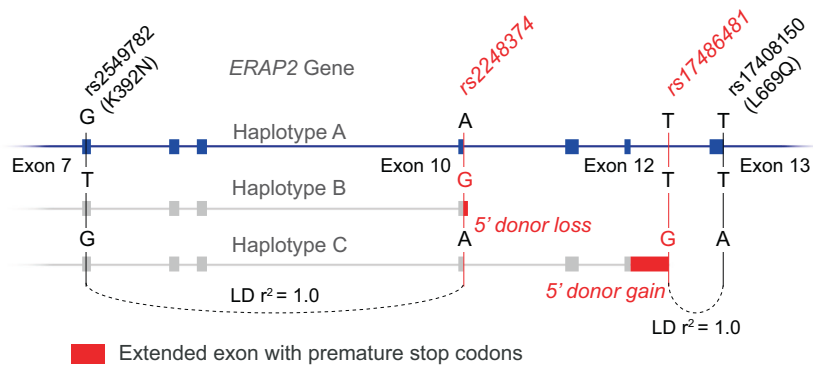


Fig. 2 Two splice region variants mediate alternative splicing of the ERAP2 gene. There are two common intronic single nucleotide variants (>1% in the global population) within ERAP2 that mediate alternative splicing into three common haplotype groups (Haplotype A-C). With variant rs2248374 (Global allele frequency = 0.55) located downstream of exon 10, the G allele disrupts a donor splice site (5' donor loss), resulting in an extended exon 10 which encodes multiple premature stop codons, resulting in mRNA decay and loss of full-length ERAP2 protein. The G allele of rs2248374 is in full linkage disequilibrium ([LD] from 1000 Genomes) with T allele of missense variant rs2549782 (392 N). A canonical splicing site motif (5' donor gain) is introduced by the G allele of rs17486481 (global allele frequency = 0.02) located in the intronic region between Exon 12 and 13. Exon 12 will be extended and include premature stop codons that are prone to nonsense mediated decay. The G allele of rs17486481 is in full LD with the A allele of missense variant rs17408150 (669Q). Haplotype A-C encode multiple naturally occurring ERAP2 protein allotypes (see Table 1).

COMMON GENETIC VARIATION CONTROLS ALTERNATIVE SPLICING OF ERAP2

In humans, alternative splicing of the 19 exons of ERAP2 is known to produce multiple transcripts that are under tight genetic control. Strong linkage disequilibrium between the SNVs in ERAP2 results in the formation of two common haplotypes often referred to as haplotype "A" and "B" [6]. These haplotypes are best distinguished by alleles of the SNV (A-to-G) rs2248374 positioned inside the canonical 5' splice site of exon 10 that by CRISPR-based mutation in genomic DNA was shown to control ERAP2 expression [20]. ERAP2 mRNA and protein are produced by constitutive splicing of transcripts from the haplotype encoding the A allele of rs2248374 [6, 21]. In contrast, the G allele of rs2248374 breaks the motif in favour of a downstream motif which extends exon 10 so it includes multiple premature termination codons, leading to degradation of steady-state mRNA through nonsense-mediated mRNA decay (NMD) (Fig. 2). Individuals homozygous for the G

allele of rs2248374 do not produce the 110 kDa ERAP2 protein, whereas individuals homozygous for the A allele express more ERAP2 protein than heterozygous individuals [6, 20–22].

Positive selection has targeted specific amino acid residues in ERAP2 during mammalian evolution [23]. The ancestral ERAP2 haplotype, found in contemporary chimpanzees, harbours the T allele of missense variant rs2549782 that encodes an asparagine at position 392 (392 N), which significantly broadens the substrate specificity of ERAP2 protein [24]. However, across modern human populations the ancestral T allele of rs2549782 is in full linkage ($r^2 = 1.0$ in 1000 Genomes) with the G allele of rs2248374 (i.e., no enzyme expression) (Fig. 2). This indicates that there was selection against 392 N during human evolution, perhaps to narrow its specificity profile because of host-pathogen interactions. Similarly, missense variant rs75263594 at position 347 downstream of the GAMEN-motif would substitute a Threonine to Methionine and may affect peptide substrate recognition away from the

Table 1. Haplotypes of common *ERAP2* missense variants and splice variants in the 1000 Genomes populations.

	SNV	rs3733905	rs80193285	rs117041256	rs75263594	rs2549782	rs34261036	rs2248374	rs17486481	rs61731306	rs17408150	rs73152140					
	Consequence	Pro214Leu	Val239Ala	Ser289Ala	Thr347Met	Lys392Asn	Leu411Arg	Splice	Splice	Gly661Val	Leu669Gln	Leu952Pro					
	Reference	C (0.97)	T (0.993)	T (0.97)	C (0.99)	G (0.44)	T (0.997)	A (0.45)	T (0.98)	G (0.994)	T (0.98)	T (0.997)					
	Alternate	T (0.03)	C (0.007)	G (0.03)	T (0.01)	T (0.55)	G (0.003)	G (0.55)	G (0.02)	T (0.006)	A (0.02)	C (0.003)	Frequency (%)				
Hap	Amino acid position	214	239	289	347	392	411	-	-	661	669	952	AFR	AMR	EAS	EUR	SAS
A	(Ancestral) ERAP2*00	P	V	S	T	N	L	A	T	G	L	L	0	0	0	0	0
A	ERAP2*01	P	V	S	T	K	L	A	T	G	L	L	37.14	34.87	33.83	40.95	31.08
A	ERAP2*02	L	V	S	T	K	L	A	T	G	L	L	0.15	0.72	5.16	0.40	8.79
A	ERAP2*03	P	V	S	T	K	L	A	T	V	L	L	2.27	0.29	0	0	0
A	ERAP2*04	P	V	S	T	K	R	A	T	G	L	L	0.15	0	0	0.20	0.92
A	ERAP2*05	P	V	A	T	K	L	A	T	G	L	L	0	1.15	8.23	1.09	4.09
B	ERAP2*06A	P	V	S	T	N	L	G	T	G	L	L	56.35	55.19	52.48	48.91	51.94
B	ERAP2*07A	P	V	S	T	N	L	G	T	G	L	P	1.13	0	0	0	0
B	ERAP2*08A	P	V	S	M	N	L	G	T	G	L	L	0.23	2.59	0	3.08	0.41
B	ERAP2*09A	P	A	S	T	N	L	G	T	G	L	L	2.42	0.14	0	0	0
C	ERAP2*10A	P	V	S	T	K	L	A	G	G	Q	L	0.15	5.04	0	5.37	2.35

The *ERAP2* haplotypes were determined by using the phased genotype data from the 5 super populations of the 1000 Genomes. Haplotypes (Hap A) predicted to encode protein allotypes are indicated by bold black amino acid (single letter code). Haplotypes predicted to be subjected to nonsense mediated mRNA decay (Hap B and Hap C) controlled by splice variants (allele associated with alternative splicing are shown in red) are indicated in grey. The suffix 'A' (for 'aberrant', in accordance with HLA nomenclature) has been given to alternatively spliced transcripts due to the possibility of conditional expression or a lack of sufficient information to determine whether they are 'null' alleles. Non-ancestral AAs are shown as orange shaded boxes. AFR African, AMR Ad Mixed American, EAS East Asian, EUR European, SAS South Asian.

N-terminus of ERAP2 (Efstratios Stratikos, Personal Communication), but the G allele of this variant is only found in haplotype B (D' with rs2248374 = 1.0), which also precludes its translation into protein (Table 1). Perhaps similar selective mechanisms apply to another common variant rs17486481 located deep into intron 12–13 (allele frequency 5% in the European [EUR] superpopulation of the 1000 Genomes). The G allele of this variant introduces a donor splice site and produces an alternatively spliced transcript [25] (which we suggest calling "Haplotype C") with an extended exon 12 that includes premature stop codons that also makes it prone to NMD. This is significant, because this splice variant is in full linkage with the A allele of missense variant rs17408150 (Leu669Gln) (Fig. 2), which is predicted to alter protein stability (global protein stability change prediction metric $\Delta\Delta G_{\text{pred}} = 1.5$, data from *ProtVar*). Consequently, mRNA transcribed from haplotypes expected to produce full-length ERAP2 (rs2248374-A), but which carry the A allele of rs17408150 (Gln669) will likely be destroyed by NMD. This is supported by the significant decrease in ERAP2 protein levels associated with the A-rs17408150 (Gln669) [26]. In summary, these two common splice region variants selectively remove transcripts that encode amino acids that significantly alter the function of the enzyme. While it remains uncertain if and what selective pressures operate against or in favour of specific ERAP2 protein isoforms, a deeper understanding of ERAP2 protein allotype diversity in the population may shed light on the evolutionary forces that shaped this gene.

ERAP2 PROTEIN ALLOTYPE DIVERSITY IN THE POPULATION

Allotype diversity between human populations may indicate potential adaptation to pathogenic species. In the 1000 Genomes

Project dataset [27], the 10 most frequent missense variants in *ERAP2* in the global population form 10 haplotypes with a frequency of greater than 1% in one of the five major continental populations (Table 1). The ERAP2 allotype produced by the ancestral haplotype (which we will refer to as *ERAP2*00*) is very rare or absent in modern populations, while the most common haplotype (*ERAP2*06A*) in the population is identical to the ancestral haplotype, but harbours the G allele of rs2248374 (i.e., precludes protein expression). There are five additional haplotypes predicted to produce enzymatically active ERAP2 protein allotypes which we will refer to as *ERAP2*01-ERAP2*05*, following previously suggested nomenclature [28], and 5 predicted to not produce full-length protein allotypes; *ERAP2*06A-ERAP2*10A* (the suffix 'A' stands for 'Aberrant' expression where there is some doubt as to whether a protein is actually expressed [i.e., NMD under steady-state conditions]). The most common ERAP2 allotype (*ERAP2*01*) differs in amino acid sequence from the ancestral allotype at position 392. Additional amino acid positions determine other allotypes, such as *ERAP2*03* (661 V) allotype that is common in Africa (2.3%) but absent in European and Asian populations. Also, *ERAP2*02* (214 L) and *ERAP2*05* (289 A) allotypes are common in Asia and rare (or absent) in other continental populations. To comprehensively map the landscape of allotype diversity in the general population, analysis of much larger sample sizes is needed (e.g., The UK Biobank includes data from 500,000 participants). Since ERAP2 is involved in a variety of human conditions [2], the polymorphic nature of the protein needs to be considered and allotype-specific study between pathogenic insult and immune response may help better understand the functional relation between ERAP2 allotypes and human disease.

KEEPING HAPLOTYPES IN BALANCE

The *ERAP2* sequence of chimpanzees and bonobos is >99% identical to that of humans, and the nucleotide sequence orthogonal to human exon 10 and the downstream splice region containing rs2248374 are identical (see Ensemble Genome Browser, accessed 31 July 2023) [29]. Intriguingly, these great apes are not polymorphic at the nucleotide position orthologous to rs2248374, and consequently lack the G allele that results in alternative splicing at exon 10 [6]. Therefore, the G allele may have been introduced after the split of our common ancestor during human evolution.

Due to the scarce fossil records available during human evolution's early stages, estimations of allelic introduction remain challenging. Also, complex human behaviours, such as migration and reticulation are difficult to address in evolutionary models [30]. Like much of present-day genetic variation in humans, the estimated time of introduction of the G allele of rs2248374 in homo sapiens traces back to a common ancestor of Neanderthals and humans (>700,000 years based on simulation-based prediction of modern human genome data) [31]. It may also be possible that *ERAP2* SNVs have been introduced through interbreeding (termed 'introgression') between archaic humans and ancestors of modern Eurasian populations (~50,000 years ago) [25] (Fig. 3). Adaptive introgression of gene alleles from Neanderthals to modern humans has been shown to be driven by host-RNA virus interactions [32], and the modelling of the *ERAP2* gene suggests Neanderthal-introgressed alleles, such as rs17409040 that is in full linkage with splice variant rs17486481 (i.e., Haplotype C could be a result of introgression) [25, 33, 34]. However, unlike variants that reside in haplotype C, the G allele of rs2248374 (i.e., haplotype B) is common in all global populations, including African populations considered devoid of archaic introgression, which indicates that rs2248374 probably arose

in an early hominid ancestor and its allelic diversity in modern humans populations is perhaps result of continuous or recurrent contact between multiple very early African lineages [30].

Multiple independent studies have shown that the haplotypes of *ERAP2* have persisted over very long periods of time as the result of long-standing balancing selection (>Million years) [6, 35–37]. The process of balancing selection does not reduce genetic variation by selecting for alleles, but rather maintains multiple common alleles to conserve beneficial diversity within a population, in particular those involved in the co-evolutionary arms race between host and pathogens [37–39]. In humans, *ERAP2* does not show signature of ancient balancing selection indicating that variants were likely introduced in the human lineage after the split from the chimpanzees [40]. However, the balancing selection signature of *ERAP2* is old and estimated at over more than a million years predating the split between humans and archaic hominins with the estimated time at which alleles have coalesced to a single common ancestor that lived up to ~1–4 Million years ago [6, 35, 36]. Evidence in support of long-term balancing selection are the very high polymorphic diversity of the *ERAP2* gene compared to the average protein-coding gene, and the maintenance of alleles of SNVs at intermediate frequency across the global human population [6, 35, 37, 38].

Alternatively spliced haplotypes of *ERAP2* are likely retained in populations either to maximise fitness or to confer an unknown adaptive benefit. Indeed, haplotype A has been associated with resistance against deadly infectious diseases [3, 4, 35]. Because haplotype A exhibits linkage to specific *ERAP1* allotypes, we must consider the functional relationship between *ERAP2* and its family member *ERAP1* in evolutionary shaping of the *ERAP2* gene. Our laboratory reported in two collaborative studies that highly active *ERAP1* allotypes harboured haplotype A of *ERAP2* less frequently

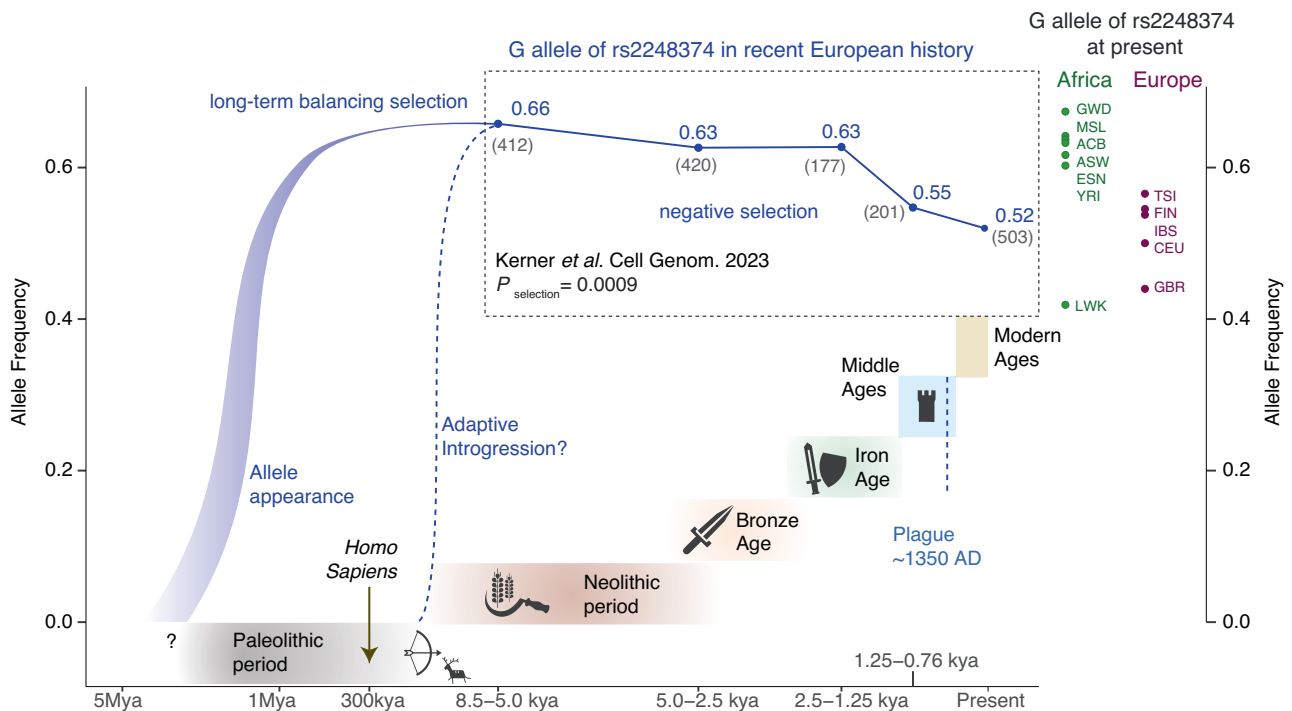


Fig. 3 Graph showing the frequency of the G allele of rs2248374 throughout human history. The dark blue line indicates the trajectory of the frequency of the G allele of rs2248374 during human evolution until present. Historic allele frequency data and selection statistics ($P_{\text{selection}}$) for rs2248374 were provided by Kerner et al. *Cell Genomics* 2023 upon request (number of ancient DNA samples used to calculate the allele frequency in grey). The allele frequency for subpopulations of the African and European populations of the 1000 Genomes Project are also shown. Kya thousand years ago, Mya million years ago, CEU Utah residents with Northern and Western European ancestry, TSI = Toscani in Italia, GBR British in England and Scotland, FIN Finnish in Finland, IBS Iberian populations in Spain, YRI Yoruba in Ibadan, Nigeria, LWK Luhya in Webuye, Kenya, GWD Gambian in Western Division, The Gambia, MSL Mende in Sierra Leone, ESN Esan in Nigeria, ASW African Ancestry in Southwest US, ACB African Caribbean in Barbados.

than intermediate active ERAP1 allotypes, while intermediate active ERAP1 allotypes had the reverse pattern [41, 42]. ERAP1 and ERAP2 can form dimers and functionally complement each other in antigen processing [1, 2]. Mapping the functional ERAP2 allotypes of patients will therefore play an important role in better understanding disease mechanisms of conditions thought to be caused by ERAP1 rather than ERAP2 [5]. Another key question is why a haplotype B persists in humans despite the loss of functional ERAP2 protein (in steady state). Work by Ye and co-workers provided tantalising clues to answer this question by demonstrating that this presumed “null allele” encodes multiple truncated ERAP2 isoforms that are selectively synthesised during influenza infection [43]. The G allele of rs2248374 has no effect on the expression of these isoforms since transcription starts at exon 9 (and lack GAMEN and HEXXH motifs essential for enzymatic activity), and includes an in-frame start site at exon 11 [43]. There is also evidence that these haplotype B-derived ERAP2 isoforms may be transcribed upon exposure to other microbial agents [44]. This has two significant implications; it proves that the widely accepted notion that a large group of individuals are “deficient” for ERAP2 may be prematurely concluded, and that alternatively spliced haplotypes may actually serve a non-redundant active immune-related purpose. Most functional research on ERAP2 is conducted in cell lines which after years of continuous cultivation are prone to profound chromosomal aberrations reported to also affect ERAP genes [45]. This emphasizes the importance of studying ERAP2 under physiological inflammatory conditions (during infection and inflammation) and in primary tissues to answer outstanding questions regarding its biological functions. Note that a common SNV in the ERAP1 gene also confers differential susceptibility to influenza infection, suggestive of similar pathogen-driven selection on the ERAP1 gene [46]. These observations support that pathogenic pressure may have enforced selection on the ERAP2 locus and that the maintenance of multiple functionally distinct haplotypes may indicate that ERAP2 function is non-redundant in humans.

RECENT GENETIC ADAPTATION OF ERAP2

While balanced alleles may be maintained for millions of years, they may change upon altering environments, such as historic migrations or changes in lifestyle [47]. In populations with profoundly different environments, such as between rainforest hunters-gatherers and farmer populations, we may expect evidence for pathogen-driven selection [48]. According to a study comparing a contemporary hunter-gatherer society with a neighbouring farmer population in Uganda that diverged more than 60,000 years ago, peripheral blood cells expressed significantly higher levels of ERAP2 after stimulating viral sensory pathways in Hunter-gatherers as compared to agriculturalists, of which the former also had higher allele frequencies of SNVs that reside in Haplotype A [49]. Coincidentally, these hunter-gatherers exhibited higher titers of serum antibodies against DNA viruses. While this makes it tempting to attribute that selection of ERAP2 variants may have been driven by differences in pathogenic exposure, the genetic selection statistics in this study did not support recent selection for ERAP2 (in line with studies of other contemporary rainforest hunter-gatherer societies [48]), which indicates that this intriguing finding was likely due to the modest sample size inherent to studies of remote populations.

However, two recent studies in European populations demonstrate that the ERAP2 alleles previously under long-term balancing selection were recently (<10,000 years ago) targeted by positive selection. A rapid positive selection for haplotype A was shown during the bubonic plague pandemic (*The Black Death*) caused by the bacterium *Yersinia pestis* that swept through Europe around 1350. Genotyping of remains of individuals before, during, and after the *Black Death* pandemic (caused by infection with *Yersinia Pestis*)

revealed significant increase in frequency of the T allele of SNV rs2549794 (tags haplotype A) over the course and after the pandemic [3]. Although the reported increase in allele frequency in comparison to other studies remains topic of debate [50, 51], macrophages from present-day individuals that carried Haplotype A neutralised *Yersinia Pestis* more effectively in an in vitro analysis [3].

A recent large cohort of 2376 ancient and 503 modern European genomes sampled over the course of 10,000 years across the Neolithic period, the Bronze Age, the Iron Age, the Middle ages, and present-day Europeans tested for allelic frequency changes in SNVs across this historical period [52]. Data from this study revealed that the G allele of rs2248374 has evolved under negative selection over the last ~3000 years (estimated time of selection onset for the G allele of rs2248374 of 2866 [95% confidence interval = 317–8,747 years], based on data kindly provided by Kerner and co-workers [52]). This observation provides evidence for short-term natural selection in favour of haplotype A in very recent European history and is conceptually in line with the observations from the analysis of genomes during the *Black Death* pandemic. These two ‘longitudinal’ studies provide supporting evidence for very recent selection changes for long-standing balanced alleles of ERAP2 as civilization spread across the European continent and encountered drastically altered environments. This local adaptation of ERAP2 haplotypes is further supported by the observation that allele frequency of the G allele of rs2248374 during the Neolithic era in Europe is very close to the allele frequency in modern African populations (~0.65), as compared to modern European populations (~0.5) (with the exception of the African *Luyha* [LWK] population, which is genetically different from other African populations [53] (Fig. 3)). While examples of genes under long-term balancing selection that show signatures of recent positive or negative selection have been described [47], more studies are needed to gain a deeper understanding of how ERAP2 evolved over the course of human evolution. Haplotype-based analysis using missense variants that define naturally occurring ERAP2 allotypes (Table 1) may aid in understanding how and to what extent the biological functions of ERAP2 adapted to environmental pressure in recent history.

ERAP2 HAPLOTYPES AND AUTOIMMUNITY

Genetic variants that reside in haplotype A of the ERAP2 gene predispose to a variety of severe inflammatory conditions, including among others, *Crohn’s* disease, ankylosing spondylitis, birdshot chorioretinopathy, and Juvenile idiopathic arthritis [5, 20]. The same alleles of autoimmune disease-risk variants have also been shown to provide increased protection against severe infections (a leading cause of death in the past) [3, 4, 52]. In a recent study of genetic data from large contemporary human biobanks, it was shown that SNV’s that tag haplotype A of ERAP2 was associated with protection against respiratory infections, with increased ERAP2 expression in whole blood (i.e., haplotype A) being associated with protection against severe respiratory infection (such as Pneumonia and Covid-19), but with opposing effects on *Crohn’s* disease [4]. This *antagonistic pleiotropy* (i.e., alleles of SNVs that exhibit both beneficial and unfavourable effects) of genetic variants in ERAP2 supports that the common haplotypes may have provided protection against infectious diseases in human history at the cost of a higher genetic risk for chronic inflammatory conditions in modern humans (Fig. 4). How haplotype A-encoded ERAP2 protein contributes to autoimmune conditions remains uncertain. Additional SNVs control ERAP2 expression by mechanisms other than alternative splicing, and variants linked to autoimmune diseases also alter enhancer-promoter interactions of ERAP2 leading to increased expression levels [20]. As a result, haplotype A may not be pathogenic in itself, but higher levels of ERAP2 associated with inflammatory processes may increase the risk of inflammatory diseases,

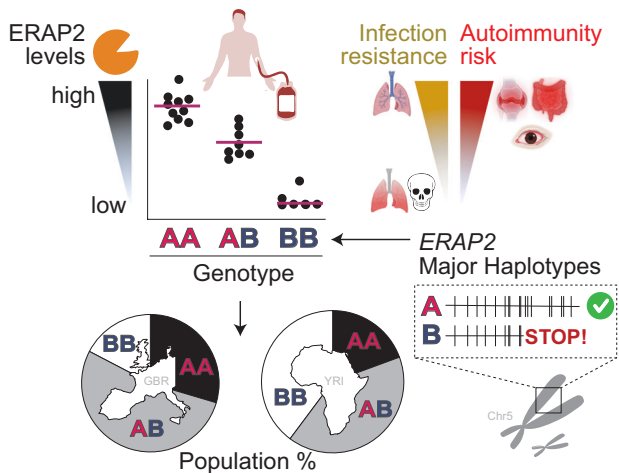


Fig. 4 The impact of ERAP2 on human health & disease. A high level of (heritable) interindividual variation in ERAP2 is caused by the two major ERAP2 haplotypes A and B, which are controlled by alternative splicing mediated by the splice variant rs2248374 (Fig. 2). Furthermore, substantial differences in ERAP2 genotype frequencies are observed between continental populations, such as between Africans and Europeans, and support genetic adaptation of ERAP2 in human history. ERAP2 enhances resistance to respiratory infections but also increases the risk of autoimmune diseases. ERAP2's role in health and disease may be better understood if we gain a deeper understanding of the function of ERAP2 protein allotype diversity across populations. The figure contains elements created with BioRender.com. GBR = British in England and Scotland, YRI = Yoruba in Ibadan, Nigeria; Genotype frequencies from GBR [= British in England and Scotland] and YRI [= Yoruba in Ibadan, Nigeria] populations of the 1000 Genomes Project.

although further studies are needed to determine whether these mechanisms are involved. The widespread adoption of single-cell transcriptomic profiling technologies will aid in capturing the biological regulation and dysfunction of ERAP2 in human pathologies. Due to the dramatic impact of common variants on ERAP2 expression levels, caution is warranted when interpreting data from studies with small sample sizes because genotype imbalances can strongly impact quantitative analyses of ERAP2 transcripts and proteins. Therefore, multi-omic (or “complementary”) single-cell technologies that also consider high-resolution genotyping in primary tissues should be used to investigate gene-by-environment interactions of ERAP2.

IN CONCLUSION

According to current data, haplotypes of ERAP2 have opposing genetic association with chronic inflammation and infection. Susceptibility to autoimmune diseases in the human population may be related to ERAP2 haplotypes that show evidence for natural selection in recent history because they are associated with resistance to infection as revealed by ancient genome studies. Further investigations of ERAP2 allotype diversity in the modern human population combined with insights from evolutionary studies that model selective pressure on alleles will enhance our understanding of its function in immunity and potentially aid in the development of targeted therapies for autoimmune conditions and infectious diseases.

DECLARATION OF 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.

DATA AVAILABILITY

The datasets used in this review are available in the 1000 Genomes Project repository <https://www.internationalgenome.org/> and allele frequencies for each population accessed via the Ensemble Genome Browser <https://www.ensembl.org/>.

REFERENCES

- Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, et al. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol.* 2005;6:689–97. <https://doi.org/10.1038/ni1208>.
- de Castro JAL, Stratikos E. Intracellular antigen processing by ERAP2: Molecular mechanism and roles in health and disease. *Hum Immunol.* 2019;80:310–7. <https://doi.org/10.1016/j.humimm.2018.11.001>.
- Klunk J, Vilgalys TP, Demeure CE, Cheng X, Shiratori M, Madej J, et al. Evolution of immune genes is associated with the Black Death. *Nature.* 2022;611:312–9. <https://doi.org/10.1038/s41586-022-05349-x>.
- Hamilton F, Mentzer AJ, Parks T, Baillie JK, Smith GD, Ghazal P, et al. Variation in ERAP2 has opposing effects on severe respiratory infection and autoimmune disease. *Am J Hum Genet.* 2023;110:691–702. <https://doi.org/10.1016/j.ajhg.2023.02.008>.
- Kuiper JJ, Prinz JC, Stratikos E, Kuśnierczyk P, Arakawa A, Springer S, et al. EULAR studygroup MHC-I-opathies. EULAR study group on ‘MHC-I-opathy’: identifying disease-overarching mechanisms across disciplines and borders. *Ann Rheum Dis.* 2023;82:887–96. <https://doi.org/10.1136/ard-2022-222852>.
- Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin SQ, Hurler B, et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genet.* 2010;6:e1001157. <https://doi.org/10.1371/journal.pgen.1001157>.
- Babaie F, Hosseinzadeh R, Ebraze M, Seyfzadeh N, Aslani S, Salimi S, et al. The roles of ERAP1 and ERAP2 in autoimmunity and cancer immunity: New insights and perspective. *Mol Immunol.* 2020;121:7–19. <https://doi.org/10.1016/j.molimm.2020.02.020>.
- López de Castro JA. How ERAP1 and ERAP2 Shape the Peptidomes of Disease-Associated MHC-I Proteins. *Front Immunol.* 2018;9:2463. <https://doi.org/10.3389/fimmu.2018.02463>.
- Tedeschi V, Paldino G, Paladini F, Matorre B, Tuosto L, Sorrentino R, et al. The Impact of the ‘Mis-Peptidome’ on HLA Class I-Mediated Diseases: Contribution of ERAP1 and ERAP2 and Effects on the Immune Response. *Int J Mol Sci.* 2020;21:9608. <https://doi.org/10.3390/ijms21249608>.
- Tsujimoto M, Hattori A. The oxytocinase subfamily of M1 aminopeptidases. *Biochim Biophys Acta.* 2005;1751:9–18. <https://doi.org/10.1016/j.bbapap.2004.09.011>.
- Evnouchidou I, Koumantou D, Nugue M, Saveanu L. M1-aminopeptidase family - beyond antigen-trimming activities. *Curr Opin Immunol.* 2023;83:102337. <https://doi.org/10.1016/j.coi.2023.102337>.
- Paladini F, Fiorillo MT, Tedeschi V, Matorre B, Sorrentino R. The Multifaceted Nature of Aminopeptidases ERAP1, ERAP2, and LNPEP: From Evolution to Disease. *Front Immunol.* 2020;11:1576. <https://doi.org/10.3389/fimmu.2020.01576>.
- Tanioka T, Hattori A, Mizutani S, Tsujimoto M. Regulation of the human leukocyte-derived arginine aminopeptidase/endoplasmic reticulum-aminopeptidase 2 gene by interferon-gamma. *FEBS J.* 2005;272:916–28. <https://doi.org/10.1111/j.1742-4658.2004.04521.x>.
- Tanioka T, Hattori A, Masuda S, Nomura Y, Nakayama H, Mizutani S, et al. Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases. *J Biol Chem.* 2003;278:32275–83. <https://doi.org/10.1074/jbc.M305076200>.
- Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet.* 2021;53:1415–24. <https://doi.org/10.1038/s41588-021-00931-x>.
- Surendran P, Feofanova EV, Lahrouchi N, Ntalla I, Karthikeyan S, Cook J, et al. The emerging multifunctional roles of ERAP1, ERAP2 and IRAP between antigen processing and renin-angiotensin system modulation. *Front Immunol.* 2022;13:1002375. <https://doi.org/10.3389/fimmu.2022.1002375>.
- Schmidt K, Leisegang M, Kloetzel PM. ERAP2 supports TCR recognition of three immunotherapy targeted tumor epitopes. *Mol Immunol.* 2023;154:61–68. <https://doi.org/10.1016/j.molimm.2022.12.010>.
- Venema WJ, Hiddingh S, de Boer JH, Claas FHJ, Mulder A, den Hollander AI, et al. ERAP2 Increases the Abundance of a Peptide Submotif Highly Selective for the Birdshot Uveitis-Associated HLA-A29. *Front Immunol.* 2021;12:634441. <https://doi.org/10.3389/fimmu.2021.634441>.

19. Kuiper JJW, Venema WJ. HLA-A29 and Birdshot Uveitis: Further Down the Rabbit Hole. *Front Immunol.* 2020;11:599558. <https://doi.org/10.3389/fimmu.2020.599558>.
20. Venema WJ, Hiddingh S, van Loosdregt J, Bowes J, Ballin B, de Boer JH, et al. A cis-regulatory element regulates ERAP2 expression through autoimmune disease risk SNPs. *Cell Genomics* 2023. Accepted for publication. <https://doi.org/10.1101/2023.03.03.530973>.
21. Coulombe-Huntington J, Lam KC, Dias C, Majewski J. Fine-scale variation and genetic determinants of alternative splicing across individuals. *PLoS Genet.* 2009;5:e1000766. <https://doi.org/10.1371/journal.pgen.1000766>.
22. Kuiper JJ, Van Setten J, Ripke S, Van T, Slot R, Mulder F, et al. A genome-wide association study identifies a functional ERAP2 haplotype associated with birdshot chorioretinopathy. *Hum Mol Genet.* 2014;23:6081–7. <https://doi.org/10.1093/hmg/ddu307>.
23. Forni D, Cagliani R, Tresoldi C, Pozzoli U, De Gioia L, Filippi G, et al. An evolutionary analysis of antigen processing and presentation across different time-scales reveals pervasive selection. *PLoS Genet.* 2014;10:e1004189. <https://doi.org/10.1371/journal.pgen.1004189>.
24. 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;189:2383–92. <https://doi.org/10.4049/jimmunol.1200918>.
25. Brand CM, Colbran LL, Capra JA. Resurrecting the alternative splicing landscape of archaic hominins using machine learning. *Nat Ecol Evol.* 2023;7:939–53. <https://doi.org/10.1038/s41559-023-02053-5>.
26. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558:73–9. <https://doi.org/10.1038/s41586-018-0175-2>.
27. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74. <https://doi.org/10.1038/nature15393>.
28. 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 USA.* 2014;111:17594–9. <https://doi.org/10.1073/pnas.1408882111>.
29. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. *Nucleic Acids Res.* 2022;50:D988–D995. <https://doi.org/10.1093/nar/gkab1049>.
30. Ragsdale AP, Weaver TD, Atkinson EG, Hoal EG, Möller M, Henn BM, et al. A weakly structured stem for human origins in Africa. *Nature.* 2023;617:755–63. <https://doi.org/10.1038/s41586-023-06055-y>.
31. Albers PK, McVean G. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS Biol.* 2020;18:e3000586. <https://doi.org/10.1371/journal.pbio.3000586>.
32. Enard D, Petrov DA. Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell.* 2018;175:360–371.e13.
33. Findley AS, Zhang X, Boye C, Lin YL, Kalita CA, Barreiro L, et al. A signature of Neanderthal introgression on molecular mechanisms of environmental responses. *PLoS Genet.* 2021;17:e1009493.
34. McCoy RC, Wakefield J, Akey JM. Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. *Cell.* 2017;168:916–927.e12.
35. Cagliani R, Riva S, Biasin M, Fumagalli M, Pozzoli U, Lo Caputo S, et al. Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. *Hum Mol Genet.* 2010;19(Dec 1):4705–14. <https://doi.org/10.1093/hmg/ddq401>.
36. Bitarello BD, de Filippo C, Teixeira JC, Schmidt JM, Kleinert P, Meyer D, et al. Signatures of Long-Term Balancing Selection in Human Genomes. *Genome Biol Evol.* 2018;10:939–55. <https://doi.org/10.1093/gbe/evy054>.
37. Cheng X, DeGiorgio M. Flexible Mixture Model Approaches That Accommodate Footprint Size Variability for Robust Detection of Balancing Selection. *Mol Biol Evol.* 2020;37:3267–91. <https://doi.org/10.1093/molbev/msaa134>.
38. Bitarello BD, Brandt DYC, Meyer D, Andrés AM. Inferring Balancing Selection From Genome-Scale Data. *Genome Biol Evol.* 2023;15:evad032. <https://doi.org/10.1093/gbe/evad032>.
39. Key FM, Teixeira JC, de Filippo C, Andrés AM. Advantageous diversity maintained by balancing selection in humans. *Curr Opin Genet Dev.* 2014;29:45–51. <https://doi.org/10.1016/j.gde.2014.08.001>.
40. Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science.* 2013;339:1578–82. <https://doi.org/10.1126/science.1234070>.
41. Kuiper JJW, Setten JV, Devall M, Cretu-Stancu M, Hiddingh S, Ophoff RA, et al. Functionally distinct ERAP1 and ERAP2 are a hallmark of HLA-A29 (Birdshot) Uveitis. *Hum Mol Genet.* 2018;27:4333–43. <https://doi.org/10.1093/hmg/ddy319>.
42. Hutchinson JP, Temponeras I, Kuiper J, Cortes A, Korczynska J, Kitchen S, et al. Common allotypes of ER aminopeptidase 1 have substrate-dependent and highly variable enzymatic properties. *J Biol Chem.* 2021;296:100443. <https://doi.org/10.1016/j.jbc.2021.100443>.
43. Ye CJ, Chen J, Villani AC, Gate RE, Subramaniam M, Bhangale T, et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. *Genome Res.* 2018;28:1812–25. <https://doi.org/10.1101/gr.240390.118>.
44. Saulle I, Vanetti C, Goglia S, Vicentini C, Tombetti E, Garziano M, et al. 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.* 2020;9:1951. <https://doi.org/10.3390/cells9091951>.
45. Heterozygosity of the 721.221-B*51:01 Cell Line Used in the Study by Guasp et (Arthritis Rheumatol, February 2016). *Arthritis Rheumatol.* Mar;69:686. <https://doi.org/10.1002/art.40073>. Erratum for: *Arthritis Rheumatol.* 2016;68:505-15.
46. Schott BH, Wang L, Zhu X, Harding AT, Ko ER, Bourgeois JS, et al. Single-cell genome-wide association reveals that a nonsynonymous variant in ERAP1 confers increased susceptibility to influenza virus. *Cell Genom.* 2022;2:100207. <https://doi.org/10.1016/j.xgen.2022.100207>.
47. de Filippo C, Key FM, Ghirotto S, Benazzo A, Meneu JR, Weihmann A, et al. Recent Selection Changes in Human Genes under Long-Term Balancing Selection. *Mol Biol Evol.* 2016;33:1435–47. <https://doi.org/10.1093/molbev/msw023>.
48. Lopez M, Choin J, Sikora M, Siddle K, Harmant C, Costa HA, et al. Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest. *Curr Biol.* 2019;29:2926–2935.e4. <https://doi.org/10.1016/j.cub.2019.07.013>.
49. Harrison GF, Sanz J, Boulais J, Mina MJ, Grenier JC, Leng Y, et al. Natural selection contributed to immunological differences between hunter-gatherers and agriculturalists. *Nat Ecol Evol.* 2019;3:1253–64. <https://doi.org/10.1038/s41559-019-0947-6>.
50. Barton AR, Santander CG, Skoglund P, Moltke I, Reich D, Mathieson I. Insufficient evidence for natural selection associated with the Black Death. *bioRxiv.* 2023. <https://doi.org/10.1101/2023.03.14.532615>.
51. Vilgalys TP, Klunk J, Demeure CE, Cheng X, Shiratori M, Madej J, et al. Reply to Barton et al: signatures of natural selection during the Black Death. *bioRxiv.* 2023. <https://doi.org/10.1101/2023.04.06.535944>.
52. Kerner G, Neehus AL, Philippot Q, Bohlen J, Rinchai D, Kerrouche N, et al. Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell Genom.* 2023;3:100248. <https://doi.org/10.1016/j.xgen.2022.100248>.
53. Sikora M, Carpenter ML, Moreno-Estrada A, Henn BM, Underhill PA, Sánchez-Quinto F, et al. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS Genet.* 2014;10:e1004353. <https://doi.org/10.1371/journal.pgen.1004353>.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 954992. We are grateful to the members Efstathios Stratikos, Piotr Kuśnierczyk, and Akiko Arakawa of the EULAR study group for 'MHC-I-opathy' that provided guidance and constructive feedback on ERAP2 allotype nomenclature. We would also like to thank Manuela Sironi and Rachele Cagliani for their insightful discussions regarding the balancing selection of ERAP2.

AUTHOR CONTRIBUTIONS

AR and JK were responsible for writing the review and making the figures.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Jonas J. W. Kuiper.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023