Variation in ERAP2 has opposing effects on severe respiratory infection and autoimmune disease

Variation in *ERAP2* has opposing effects on severe respiratory infection and autoimmune disease

Authors

Fergus Hamilton, Alexander J. Mentzer, Tom Parks, J Kenneth Baillie, George Davey Smith, Peter Ghazal, Nicholas J. Timpson

Correspondence

fergus.hamilton@bristol.ac.uk

The T allele at the *ERAP2* SNP rs2549794 has been suggested as deleterious during the Black Death. We have found that the T allele is associated with increased risk of respiratory infection, with opposing effects for Crohn disease, supporting the hypothesis of balancing selection driven by autoimune and infectious disease.

Hamilton et al., 2023, The American Journal of Human Genetics *110*, 1–12
April 6, 2023

https://doi.org/10.1016/j.ajhg.2023.02.008

CellPress

Please cite this article in press as: Hamilton et al., Variation in *ERAP2* has opposing effects on severe respiratory infection and autoimmune disease, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.02.008

REPORT

Variation in *ERAP2* has opposing effects on severe respiratory infection and autoimmune disease

Fergus Hamilton,[1,2,*] Alexander J. Mentzer,[4] Tom Parks,[4,5] J Kenneth Baillie,[6,7,8] George Davey Smith,[1] Peter Ghazal,[9] and Nicholas J. Timpson[1,3]

Summary

ERAP2 is an aminopeptidase involved in immunological antigen presentation. Genotype data in human samples from before and after the Black Death, an epidemic due to *Yersinia pestis*, have marked changes in allele frequency of the single-nucleotide polymorphism (SNP) rs2549794, with the T allele suggested to be deleterious during this period, while *ERAP2* is also implicated in autoimmune diseases. This study explored the association between variation at *ERAP2* and (1) infection, (2) autoimmune disease, and (3) parental longevity. Genome-wide association studies (GWASs) of these outcomes were identified in contemporary cohorts (UK Biobank, FinnGen, and GenOMICC). Effect estimates were extracted for rs2549794 and rs2248374, a haplotype tagging SNP. Additionally, *cis* expression and protein quantitative trait loci (QTLs) for *ERAP2* were used in Mendelian randomization (MR) analyses. Consistent with decreased survival in the Black Death, the T allele of rs2549794 showed evidence of association with respiratory infection (odds ratio; OR for pneumonia 1.03; 95% CI 1.01–1.05). Effect estimates were larger for more severe phenotypes (OR for critical care admission with pneumonia 1.08; 95% CI 1.02–1.14). In contrast, opposing effects were identified for Crohn disease (OR 0.86; 95% CI 0.82–0.90). This allele was shown to associate with decreased *ERAP2* expression and protein levels, independent of haplotype. MR analyses suggest that *ERAP2* expression may be mediating disease associations. Decreased *ERAP2* expression is associated with severe respiratory infection with an opposing association with autoimmune diseases. These data support the hypothesis of balancing selection at this locus driven by autoimmune and infectious disease.

Introduction

The ER aminopeptidases *ERAP1* (MIM: 606832) and *ERAP2* (MIM: 609497) both code for aminopeptidases that are critical in presentation of antigen by professional antigen-presenting cells and are linked with the human leukocyte antigen (HLA) response to infection.[2–5] Variation at these genes has been associated with both infection and autoimmune disease, with compelling associations reported between variation at these genes and birdshot uveitis[6] (an eye disorder [MIM: 605808]), inflammatory bowel disease (particularly Crohn disease [MIM: 266600]), Behcet's disease[7] (MIM: 109650), and ankylosing spondylitis (MIM 106300).[8] Other studies have suggested a role for an association with *ERAP2* and infection, although studies have been small. Additionally, it is thought that these genes are under balancing selection, with *ERAP2* having two, almost equally common haplotypes worldwide (haplotype A and haplotype B).[9–11] Haplotype A encodes the full ERAP2 protein, whereas haplotype B has a premature stop codon leading to nonsense-mediated decay and reduced levels of ERAP2 protein.[12] Around 50% of all variation in *ERAP2* whole blood expression is down to haplotype, although other genetic variation near *ERAP2* appears to contribute significantly to measured expression levels.[9,13]

A recent analysis of human genomes from before, during, and after the Black Death in Europe, an epidemic caused by *Yersinia pestis*, identified a large change in allele frequency over this period in a common SNP (rs2549794-T) in *ERAP2* with targeting sequencing of immune genes.[12] This SNP is in linkage disequilibrium (LD) with the putative splicing variant (rs2248374-G) that leads to transcription of the specific haplotype (haplotype B) that is associated with production of a truncated ERAP2 protein which undergoes nonsense-mediated decay and appears to affect antigen presentation.[9]

This new work has suggested that rs2549794 is associated with the immune response, with carriers of the putative protective C allele having a 5-fold higher *ERAP2* expression in both unstimulated and *Y. pestis*-challenged macrophages, altered gene expression patterns in multiple other immune cells, and increased production of the full-length ERAP2 protein in macrophages. Additionally, the rs2549794-C allele has been associated with increased odds of Crohn disease in other studies, although it is unknown whether this is the causal variant.[14]

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; [2]Infection Science, North Bristol NHS Trust, Bristol, UK; [3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; [4]Wellcome Centre For Human Genetics, University of Oxford, Oxford, UK; [5]Department of Infectious Disease, Imperial College London, London, UK; [6]Baillie Gifford Pandemic Science Hub, Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK; [7]Roslin Institute, University of Edinburgh, Edinburgh, UK; [8]Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, UK; [9]Project Sepsis, Cardiff University, Cardiff, UK
*Correspondence: fergus.hamilton@bristol.ac.uk
https://doi.org/10.1016/j.ajhg.2023.02.008.

Please cite this article in press as: Hamilton et al., Variation in *ERAP2* has opposing effects on severe respiratory infection and autoimmune disease, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.02.008

Therefore, the study concludes that variation at *ERAP2* (in particular, the C allele at rs2549794) is probably associated with protection from *Y. pestis* and suggests that the balancing selection at this locus is likely driven by a balance between autoimmune disease (or other negative consequences of this SNP) and protection from severe infection. This conclusion remains a hypothesis and there is little data on whether variation at *ERAP2* associates with severe infection and other autoimmune diseases outside Crohn's in contemporary datasets.[15,16] It also remains unknown whether the associations at rs2549794 represent simply associations related to LD to the known haplotype structure, or whether there is additional relevant genetic variation at this locus.

In this study, we aim to test this hypothesis by identifying whether variation at *ERAP2* is associated with serious infection (sepsis and respiratory infection), autoimmune disease, and age at parental death (as a proxy for potential effects on longevity) using the two SNPs identified as associated with disease phenotypes in recent literature (rs2549794-C, identified recently, and rs2248374-G, which tags the haplotype) and by using expression[17,18] and protein[19] quantitative trait loci for *ERAP2* in Mendelian randomization analyses.

Material and methods

Study design

For this study, we aimed to assess whether variation at *ERAP2* was associated with severe infection, autoimmune disease, and parental longevity, as a proxy for an effect on lifespan.

Infection outcomes

For our pneumonia and sepsis outcomes, we utilized our previously performed case-control GWAS of sepsis and pneumonia from UK Biobank.[20] UK Biobank is a large (n $\sim$ 500,000) cohort of older (>50 years old) UK residents who were recruited between 2002 and 2009 and who have linked electronic health record and mortality data. Details of the cohort are available elsewhere.[21]

For this study we included data on both incidence outcomes (presence of disease), mortality outcomes (death within 28 days of disease), and critical-care-related outcomes (critical care admission with the disease).

Affected individuals were identified by the presence of ICD-10 coding in linked hospital data and deaths from nationally linked mortality data.[21] Comparisons were performed against all other participants in UK Biobank (e.g., death within 28 days of pneumonia vs. all other participants in UK Biobank). All analyses were performed in participants of European ancestry (n $\sim$ 350,000). Further details on case definition, analysis approach, and GWAS methodology are with the original publication[21] and elsewhere,[22] while GWASs are available at the MRC-IEU GWAS database.[1]

Additionally, we identified matching GWASs for disease incidence from the FinnGen Round 7.[23] FinnGen is a similar large population cohort in Finland, with linked genetic and health outcome data. Case-control GWASs have been performed for ICD-10-coded hospital admission, with details available at the FinnGen website.[24]

Finally, we identified COVID-19 outcomes from the GenoMICC[25] (Round 3) consortium. GenoMICC is available as an open-source international study recruiting critically ill individuals starting in 2015, with this analysis focusing on the comparison between hospitalized individuals with COVID-19 and the general population (hospitalized COVID-19) and comparisons between severely unwell individuals with COVID-19 and the general population (critical COVID-19).

Autoimmune disease outcomes

We focused on four common autoimmune diseases which have existing GWAS summary statistics available: type 1 diabetes, asthma, Crohn disease, and rheumatoid arthritis. For type 1 diabetes, we utilized summary statistics from a recent meta-analysis of 12 European cohorts.[25] For asthma, we utilized a recent GWAS in UK Biobank using a broad definition of asthma.[26] For Crohn disease, we utilized a GWAS meta-analysis performed in a European population from the International Inflammatory Bowel Disease Genetics Consortium.[14] Finally, for rheumatoid arthritis, we utilized summary statistics from a recent meta-analysis of rheumatoid arthritis.[9]

Parental longevity outcomes

Parental longevity was chosen as a proxy measurement to identify potential effects of exposures on lifespan. For this analysis, we utilized GWAS summary data from a recent GWAS by Pilling et al. in UK Biobank.[21] On recruitment to UK Biobank, participants were asked what age both parents had died at, and these data were used to generate age at death GWAS. Exclusion criteria were applied to exclude extremely early death, with details in the manuscript. To generate the combined across parents age at death data, each age was converted into a Z score, and then this was summed to generate a combined phenotype.

Individual SNP analysis

We had an iterative analytic approach. Firstly, we focused on the SNP identified by Klunk et al.[22] (rs2549794) and identified an extracted association estimate from the outcome GWAS. Secondly, as ERAP2 is characterized by two equally common haplotypes (A + B), we extracted effect estimates for the tagging SNP rs2248374, where the G allele is known to associate with haplotype B.[27]

Conditional analysis and Mendelian randomization

As the effect estimates were larger at rs2549794 than at the haplotype tagging SNP, we went on to perform a set of conditional analyses and Mendelian randomization analyses to determine whether this was related to gene expression at this locus. We identified summary statistics for expression and protein quantitative trait loci (QTLs) *cis* to *ERAP2*. Expression QTLs were identified from whole blood (from eQTLGen,[28] n = 31,864) and eQTLs from lung tissue from the GTEx consortium (n = 515).[17] Protein QTLs were identified from GWASs performed by the DECODE consortium (n = 33,559). For whole blood and protein levels, all *cis* SNPs were identified, while only significant ($p < 5 \times 10^{-8}$) SNPs were available for lung tissue.

We then used GCTA (v.1.94) to identify the effect of rs2549794 on gene expression, protein levels, and each outcome sequentially, conditioning on haplotype (by using the tagging SNP rs2248374). We used UK Biobank to provide genotypes with which to construct the LD matrix.[29] This analysis generates conditional estimates and standard errors.

Subsequently, we performed Mendelian randomization (MR). MR uses SNPs as instrumental variables to explore the association between an exposure (e.g., *ERAP2* expression) and an outcome (e.g., pneumonia).[30] Mendelian randomization is a form of instrumental variable analysis, which under certain assumptions can provide causal estimates of the effect of an exposure on an outcome. These assumptions are that the genetic instruments are associated with the risk factor of interest, were independent of potential confounders, and could only affect the outcome through the risk factor and not through alternative pathways (that is, through pleiotropy).[30] In this study, our instruments were all *cis* acting QTLs.

For all of these exposures, SNPs robustly ($p < 5 \times 10^{-8}$) associated with the exposure (expression or protein levels) within 300 kb of *ERAP2* were extracted. SNPs were pruned to ensure independence ($r^2 = 0.1$). Effect alleles were harmonized with the outcome datasets. MR was then performed for each SNP individually and results meta-analyzed via fixed effects inverse variance weighting (IVW). Other meta-analytical methods are reported as sensitivity analyses.

As a further sensitivity analysis, we performed a specific analysis where rs2248374, which tags the haplotype, was dropped from the IVW meta-analysis. This has the effect of removing this large effect from the analysis and essentially provides an MR estimate independent of ERAP2 haplotype. To explore and identify heterogeneity in SNPs, and to determine whether rs2248374 was an outlier, we used radial MR and generated IVW radial plots.[31]

Additionally, to confirm independence of our SNPSs, we performed stepwise model selection using GCTA to identify conditionally independent SNPs and used these SNPs within MR.[21] The p value threshold to identify conditional independence was 0.0001. To further test for independence, we pruned the SNPs further using a r^2 threshold of <0.01 (for whole blood expression only) and re-ran analyses.

Software

Analysis was performed using R v.4.0.4 (R Foundation for Statistical Computing, Vienna). Data wrangling was performed using the tidyverse and plotting using ggplot2.[32] Two-sample Mendelian randomization was performed using the TwoSampleMR package in R.[1] GCTA (v.1.94) was used to perform conditional analyses.

Guidelines

This study is reported in line with the STROBE-MR guidance (Document S2).[33]

Ethics

This work uses mostly publicly available data. To generate conditional estimates, we utilized individual genotype data from UK Biobank (application number 56243). UK Biobank was approved by the North West REC.

Results

We identified genome-wide association study (GWAS) summary statistics for 7 respiratory infection outcomes, 5 sepsis outcomes, 4 autoimmune disease outcomes, and 3 (mother, father, both) parental longevity-related outcomes. Table 1 lists details of each study, with descriptions of the cohorts in the material and methods.

Associations at rs2549794

In our initial analysis we focused on rs2549794 as a potentially causal SNP with some functional evidence of an effect on cellular phenotypes.[35] This SNP has a minor allele frequency of between 0.4 and 0.5 in modern European ancestry cohorts (MAF 0.40 in 1000 Genomes, 0.41 in gnomAD). The T allele of rs2549794 was found to be associated with increased odds of respiratory infection (Figure 1) but not sepsis. Effect sizes increased with severity of disease, with the largest effects for death from critical care pneumonia (OR 1.11; 95% CI 0.98–1.26, $p = 0.094$) and for critical care admission with pneumonia (OR 1.09; 95% CI 1.02–1.14, $p = 0.008$).

The effect size on likely bacterial respiratory infection was larger than that on viral infection (COVID-19 in GenoMICC),[23] although all effects were modest in size (OR < 1.15). As expected, this allele was associated with protection from Crohn disease (OR 0.86; 95% CI 0.82–0.90, $p = 8.6 \times 10^{-9}$) and we identified a protective association with type 1 diabetes (MIM: 222100) (OR 0.95; 95% CI 0.90–0.99, $p = 0.02$, all studies in Table 1). In this dataset, we were unable to resolve any association with asthma (MIM: 600807) or rheumatoid arthritis (MIM: 180300). There was evidence of an association with parental longevity, with each additional T allele being associated with increased age at parental death (beta 0.01 on Z-scored combined age of death of parents, 95% CI 0.004–0.017, $p = 0.002$), with no difference between maternal and paternal estimates. Table 2 shows results of this analysis.

Association with ERAP2 haplotype and gene expression

As rs2549794 is in linkage disequilibrium with rs2248374 (LD between 0.6 and 0.8 in European populations), which tags the haplotype and is a splicing variant, and *ERAP2* haplotype has large effects on gene expression and disease,[9,11] we then went on to look at the effect of *ERAP2* haplotype on these outcomes, using rs2248374-G as a tagging SNP to represent haplotype B, which is associated with reduced gene expression of *ERAP2*. These results are reported in Figure 2 and Table S1. The effect estimates were broadly similar to those at rs2549794, consistent with the LD structure at this locus, although effect estimates were generally larger in rs2549794 than using the haplotype-tagging SNP.

To explore this change in estimate further, and because previous work has suggested that variation (including the haplotype effect) at *ERAP2* acts via alternative splicing leading to altered RNA and protein levels,[2,9,12] we then identified datasets with available expression (whole blood) and protein quantitative trait loci *cis* to ERAP2, to enable us to perform conditional analysis and Mendelian randomization. Summary statistics for expression associations were accessed from the eQTLGen consortium while protein associations were accessed from the DECODE consortium.

Using GCTA,[36] we then performed conditional analysis to identify whether rs2549794 had any additional effect

Please cite this article in press as: Hamilton et al., Variation in *ERAP2* has opposing effects on severe respiratory infection and autoimmune disease, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.02.008

Table 1. Studies included in this analysis

Outcome	Author	Affected individuals	Control subjects	Cohort	Case definition
Respiratory infection					
Pneumonia	Hamilton et al.[34]	22,567	463,917	UK Biobank	ICD-10 coded pneumonia
Pneumonia (critical care)	Hamilton et al.[34]	2,758	428,607	UK Biobank	ICD-10 coded pneumonia and critical care admission
Pneumonia (death)	Hamilton et al.[34]	3,185	430,820	UK Biobank	ICD-10 coded pneumonia and death within 28 days
Pneumonia (death in critical care)	Hamilton et al.[34]	545	430,820	UK Biobank	ICD-10 coded pneumonia, critical care admission, and death within 28 days
Pneumonia (FinnGen)	FinnGen consortium[35]	38,999	282,303	FinnGen	ICD-10 coded pneumonia
Critical COVID	GenOMICC consortium[23]	13,327	1,745,220	meta-analysis	very severe respiratory COVID-19
Hospitalized COVID	GenOMICC consortium[23]	43,308	2,918,102	meta-analysis	hospitalized with COVID-19
Sepsis					
Sepsis (death in critical care)	Hamilton et al.[34]	347	431,018	UK Biobank	ICD-10 coded sepsis
Sepsis	Hamilton et al.[34]	11,643	474,841	UK Biobank	ICD-10 coded sepsis and critical care admission
Sepsis (critical care)	Hamilton et al.[34]	1,380	429,985	UK Biobank	ICD-10 coded sepsis and death within 28 days
Sepsis (death)	Hamilton et al.[34]	1,896	484,588	UK Biobank	ICD-10 coded sepsis, critical care admission, and death within 28 days
Sepsis (FinnGen)	FinnGen consortium[35]	10,337	310,965	FinnGen	ICD-10 coded sepsis
Autoimmune disease					
Crohn disease	Liu et al.[14]	5,956	14,927	meta-analysis	histologically and endoscopically confirmed
Type 1 diabetes	Forgetta et al.[24]	4,329	9,543	meta-analysis	cohort specific
Asthma	Valette et al.[25]	56,167	352,255	UK Biobank	ICD coded and self-reported
Rheumatoid arthritis	Ha et al.[26]	14,361	43,923	meta-analysis	cohort specific
Parental longevity					
Age at death (father)	Pilling et al.[27]	317,652	N/A	UK Biobank	age at death of father
Age at death (mother)	Pilling et al.[27]	246,941	N/A	UK Biobank	age at death of mother
Age at death (both)	Pilling et al.[27]	389,166	N/A	UK Biobank	summed *Z* score for age of death and mother

on gene expression/protein levels after conditioning on *ERAP2* haplotype, modeled by rs2248374. Despite the LD between them ($R^2 = 0.74$, $D' = 1$ in 1000 Genomes Europe), there was a clear effect on gene expression even after conditioning on haplotype (Table 3), with each T allele still associated with reduced gene expression and translation at *ERAP2*.

We then used COJO separately on our outcome GWAS, to generate estimates of the association at rs2549794 conditional on haplotype (Figure S1).[37] Associations remained in the same direction as in our primary analysis, although all estimates crossed the null, suggesting a partial but not total attenuation of effect once taking account of the effect of haplotype.

Given the stronger association at rs2549794 than the haplotype-tagging SNP rs2248374 and similar direction (although weaker) conditional estimates—and with both putatively harmful alleles associated with reduced gene expression at ERAP2—we went on to perform Mendelian randomization (MR) at ERAP2, aiming to estimate the effect of changes in gene expression and protein levels to identify whether gene expression or translation were driving our identified association.

Mendelian randomization

To perform MR, we identified independent ($r^2 > 0.1$, in the European population of 1000 Genomes Project[36]) *cis* SNPs that were associated ($p < 5 \times 10^{-8}$) with (1) whole blood gene expression, (2) protein levels, and (3) lung tissue expression of *ERAP2,* where a list of associated *cis* SNPs was available via GTEx but no full summary statistics were available. SNPs were defined as *cis* if they were within 300 kb either side of the gene body. We identified 67 SNPs associated with *ERAP2* whole blood expression (from eQTLGen,[28] n = 31,864), 80 SNPs associated with ERAP2 protein levels(from DECODE,[19] n = 33,559), and 6 SNPs associated with *ERAP2* lung tissue expression (from GTEx,[17] n = 551). Details of SNP identification and

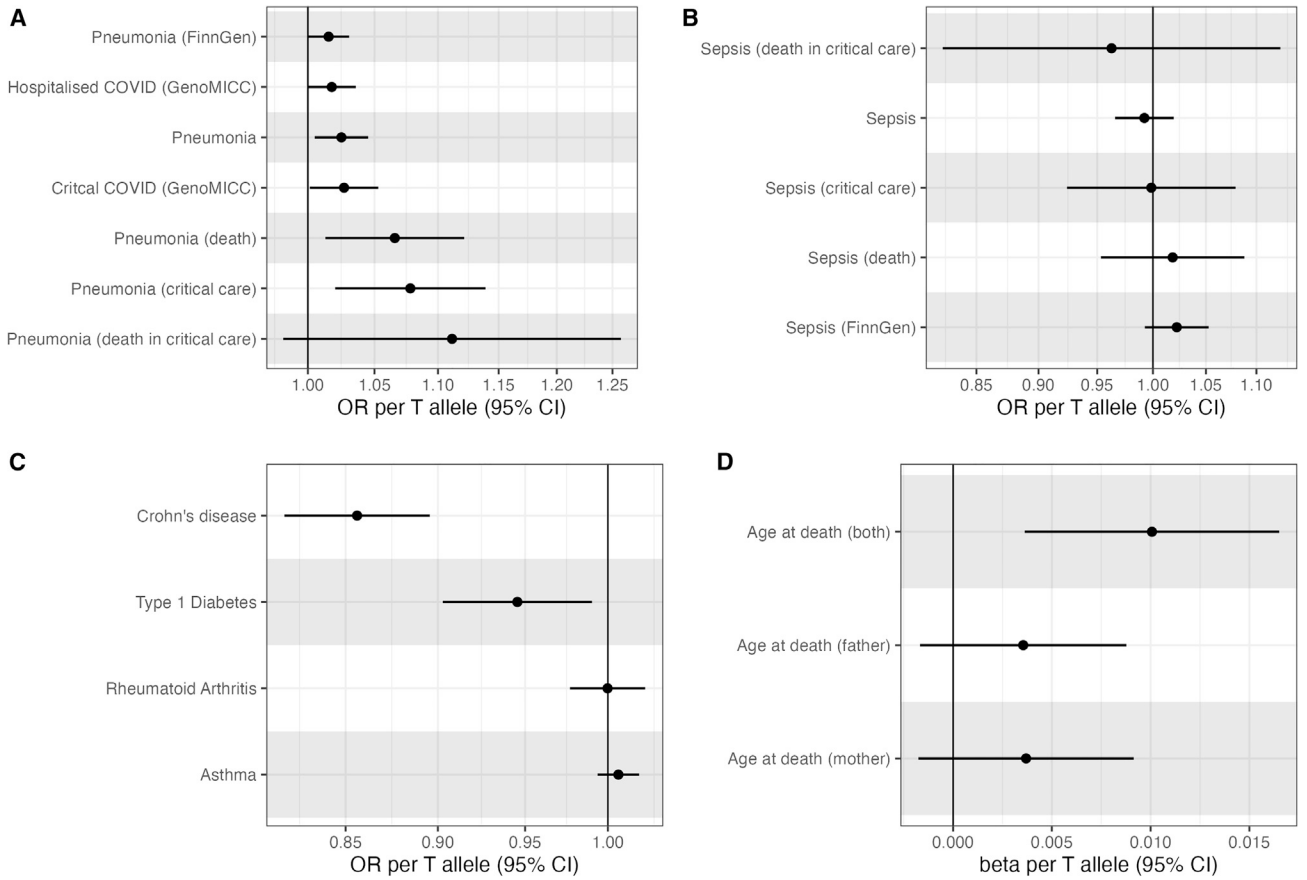

Figure 1. Associations between rs2549794 and outcomes in each GWAS dataset
Odds ratio for each outcome presented, with 95% confidence intervals, except for parental longevity, where the beta is presented with a 95% confidence interval. Outcomes: (A) respiratory infection, (B) sepsis, (C) autoimmune disease, (D) parental longevity.

pruning are in the material and methods, with a list of all included SNPs in Table S2.

As we test 19 traits with some degree of overlap between both exposures and the outcomes, we report p values in the text, but also provide p values corrected via Bonferroni (n = 19) in Tables S1–S9 where all results are reported.

We then went on to perform MR using each of these instruments independently, and we meta-analyzed each individual SNP association via inverse-variance weighted (IVW) meta-analysis to generate a summary causal estimate. In IVW meta-analysis, increased whole blood *ERAP2* expression was associated with protection from respiratory infection (Figure 3). This effect was most pronounced for the most severe infection, with the largest effect estimates for patients who died or were critically unwell with pneumonia: OR for death from critical care pneumonia 0.91; 95% CI 0.87–0.97, $p = 1.3 \times 10^{-3}$, OR for death from pneumonia 0.95; 95% CI 0.92–0.97, $p = 1.4 \times 10^{-6}$; OR for pneumonia 0.99; 95% CI 0.98–0.99, $p = 1.9 \times 10^{-3}$).

No effect was identified on sepsis-related outcomes, and opposing effects were identified on type 1 diabetes (OR 1.05; 95% CI 1.03–1.08, $p = 4.9 \times 10^{-4}$) and Crohn disease (OR 1.10, 95% CI 1.07–1.11; $p = 2.16 \times 10^{-18}$). Increased *ERAP2* expression was associated with decreased age at death in both men and women, with a combined estimate of −0.008; 95% CI −0.005 to −0.012; $p = 2.66 \times 10^{-5}$).

Using lung-tissue-specific eQTLs, effect sizes were similar, although the strength of the association was weaker, reflecting in part the fewer number of SNPs (6 SNPs vs. 67 SNPs) in these analyses (Figure S2), reducing the precision of the meta-analysis. Analyses using protein quantitative trait loci were again similar (Figure S3), although associations with COVID were null and other associations are weaker. All IVW MR QTL results for our primary instrument are reported in Tables S3–S5, while Table S6 reports alternative meta-analytic approaches (weighted median and MR Egger). These results were similar but more imprecise than our primary analysis.

Although estimates from MR with independent instruments should not be biased by background genetic structure unless it is coincident with the exposure and/or the outcome, we went on to perform a series of sensitivity analyses given the known strong association between haplotype, expression, and protein levels at this locus. Firstly, we utilized a more stringent cut off for independence ($R^2 < 0.01$) for our whole blood expression analysis to ensure that included SNPs were truly independent. This

Table 2. Associations between rs2549794 T allele and outcomes in each GWAS dataset

Outcome	OR/beta	Lower 95% CI	Upper 95% CI	p value	EAF
Respiratory infection (OR reported)					
Pneumonia	1.025	1.005	1.045	0.014	0.565
Pneumonia (critical care)	1.078	1.020	1.139	0.008	0.567
Pneumonia (death)	1.066	1.013	1.121	0.014	0.565
Pneumonia (death in critical care)	1.111	0.982	1.258	0.094	0.567
Pneumonia (FinnGen)	1.015	1.000	1.031	0.048	0.621
Critical COVID (GenoMICC)	1.027	1.001	1.053	0.038	NR
Hospitalized COVID (GenoMICC)	1.018	1.000	1.036	0.053	NR
Sepsis (OR reported)					
Sepsis (death in critical care)	0.963	0.824	1.125	0.631	0.567
Sepsis	0.992	0.966	1.019	0.567	0.565
Sepsis (critical care)	0.999	0.924	1.079	0.970	0.567
Sepsis (FinnGen)	1.022	0.993	1.053	0.144	0.622
Sepsis (death)	1.018	0.953	1.088	0.590	0.565
Autoimmune disease (OR reported)					
Asthma	1.001	0.999	1.002	0.253	0.562
Type 1 diabetes	0.945	0.903	0.990	0.017	0.570
Rheumatoid arthritis	0.990	0.968	1.012	0.380	NA
Crohn disease	0.856	0.818	0.896	8.6×10^{-9}	0.588
Age at death (beta reported)					
Age at death (father)	0.004	−0.002	0.009	0.170	0.567
Age at death (mother)	0.004	−0.002	0.009	0.190	0.567
Age at death (both)	0.010	0.004	0.0017	0.002	0.567

Odds ratio for each outcome presented, with 95% confidence intervals, except for parental longevity, where the raw beta is presented with a 95% confidence interval. NR, not reported; EAF, effect allele frequency.

analysis is reported in Table S7 and shows similar effect estimates, with reduced precision consistent with the reduction in power with fewer available SNPs (12 SNPs vs. 67 SNPs). Secondly, we removed rs2248374 from our analysis and re-ran inverse variance weighted MR which, given all other SNPs are independent, essentially removes haplotype from the analysis. Again, effects were consistent with the removal of rs2248374, accepting the expected drop in precision (Figure S4, Table S8). The correlation between the effect estimates with and without rs2248374 was 0.99.

As another test to determine whether the haplotype-tagging SNP was quantitatively different from other *cis* expression or protein QTLs, we generated radial scatterplot visualizations of effects using radial MR and used these to identify outliers (both visually and mathematically). In this analysis, the effect of each SNP on both the exposure and outcome are plotted on a radial plot in order to observe whether any individual SNP is an outlier. Two representative plots for whole blood expression and death from pneumonia and for whole blood expression and Crohn disease are shown in Figures S5A and S5B. This plot shows that rs2248374, although having a large effect, is not heterogeneous compared to other SNPs in these analyses and is not an outlier. In only 3/42 exposure-outcome associations was rs2248374 considered an outlier and in none of those cases did removal of outliers fundamentally change the IVW MR estimates. All outlier-adjusted results are in Table S9.

Finally, we generated a second set of instruments for expression and protein levels using conditional stepwise selection in GCTA,[9–12] with a p value threshold of 0.0001. This analysis identifies SNPs that are conditionally independently associated with the trait (gene expression and protein levels) and is an additional methodology to identify independent exposures to use in MR. This analysis yielded similar highly similar estimates to our estimates generated using our primary instrument, with in most cases, statistically more confident estimates (Figures S6 and S7).

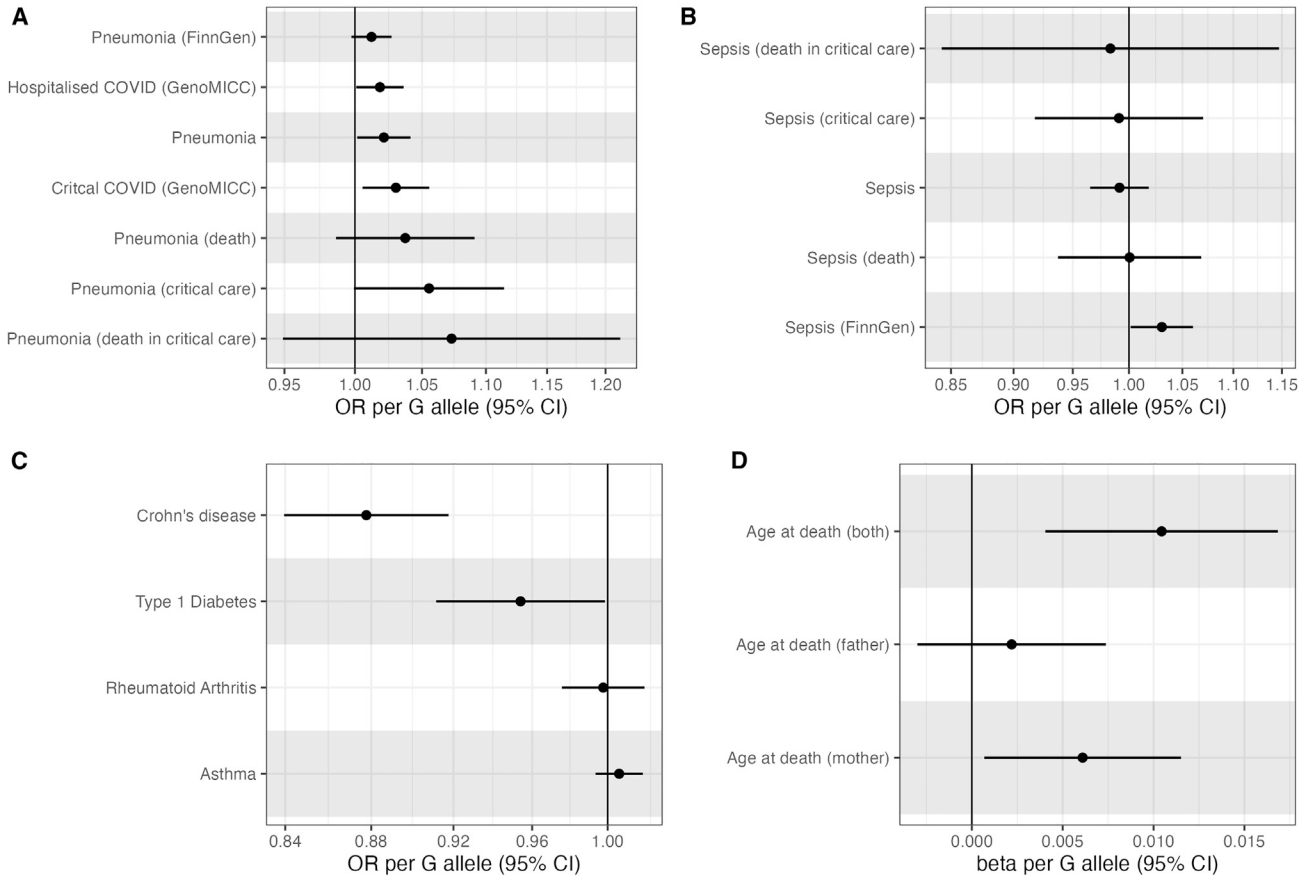

Figure 2. Association between rs2248374-G (representing haplotype B) and outcomes in each GWAS dataset
Outcomes: (A) respiratory infection, (B) sepsis, (C) autoimmune disease, (D) parental longevity.

Discussion

In this study, we show that variation at *ERAP2* is associated with susceptibility to respiratory infection in the present day. In line with a recent study using ancient DNA,[9,12] we show evidence suggesting that the C allele of rs2549794 is protective against respiratory infection, with evidence of increasing protection against severe disease. Although the effect size of the T allele on respiratory infection is modest (OR 1.05–1.1), this plausibly represents a large attenuation of the "wild-type" effect of this locus, given modern treatments for respiratory infection (e.g., antimicrobials) and other public health measures that are likely to blunt the genetic effects. In contrast, we identified an opposing effect on Crohn disease and type 1 diabetes, providing some support for the hypothesis of balancing selection at this locus.[9–12] Using parental death data from UK Biobank, we identified evidence of reduced parental lifespan with carriage of the C allele at rs2549794.

ERAP2 is characterized by two haplotypes (A + B), with haplotype B leading to reduced expression of *ERAP2* and low amounts of truncated ERAP2 protein.[5,28] This reduced amount and function of ERAP2 alters the diversity of antigens presented by HLA, and is plausibly likely to reduce the quality of immune response to certain pathogens.[5,38] As rs2549794 is in a degree of LD with the haplotype in European populations, we went on to assess whether haplotype alone was associated with the outcomes. Using a tagging SNP for haplotype B (rs2248374-G), we show a similar but reduced effect, with effect sizes marginally smaller. To explore this further, we performed conditional analyses and showed that the T allele of rs2549794 was associated with reduced expression and protein levels of ERAP2, independently of background haplotype. As rs2549794-T is in LD with rs2248374-G, with both variants associated with reduced gene expression, this suggests that gene expression may be mediating the associations we identify.

Therefore, we went on to perform Mendelian randomization analyses using multiple, independent expression and protein quantitative trait loci at *ERAP2*. These results showed clear evidence of reduced odds of severe respiratory infection with increasing levels of *ERAP2* expression or protein, but increased odds of Crohn disease and type 1 diabetes. The evidence of association was strongest with whole blood eQTL data, although we saw effects using all three of our instruments (whole blood eQTL, lung eQTL, plasma pQTL). Importantly, although the haplotype-tagging SNP provided a large effect, removal of this SNP from the analyses made little difference to estimates and there was no evidence that this had a qualitatively

Please cite this article in press as: Hamilton et al., Variation in *ERAP2* has opposing effects on severe respiratory infection and autoimmune disease, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.02.008

Table 3. Effect of the T allele of rs2549794 on *ERAP2* expression/protein levels

Quantitative trait loci	Raw beta	Raw SE	Raw p value	Conditional beta	Conditional SE	Conditional p value
Whole blood gene expression	−1.129	0.008	$<1 \times 10^{-320}$	−0.169	0.00845344	1.50×10^{-88}
Protein levels	−1.050	0.003	$<1 \times 10^{-320}$	−0.431	0.03234080	1.42×10^{-40}

different effect to other SNPs. These results suggested that variation at *ERAP2* outside the known haplotypic variation is associated with both severe infection and auto immune disease and suggest that gene expression (independent of haplotype) is causal here.

We saw no strong evidence for association with data based on sepsis-related clinical coding. This is interesting, as most data point to the bubonic form of *Y. pestis* as the predominant feature of the Black Death, although the speed of travel of the epidemic has made some consider whether pneumonic plague was a more plausible candidate.[38] It is also important to recognize that sepsis represents a dysregulated response to infection[39] and is highly heterogeneous over even the last 50 years, perhaps explaining the lack of association.[40] Although it has been long suspected that *ERAP2* has a role in protection from infection (reviewed recently[40]), with known roles in both antigen presentation and shaping the cytokine response,[41] there are no studies as far as we are aware that have robustly identified genetic associations on the population level. There are, however, some candidate gene studies reported, although given the small sample sizes their relevance to our analyses is unclear.[2,3,10,16,42]

We identified an association with reduced parental longevity with the C allele and with increased *ERAP2* expression. It is possible to speculate about the cause of this, and what the role of selection has been at this locus. One potential explanation is environmental- and/or time-specific selection, whereby in times of high pathogen pressure there is strong evolutionary pressure to increase gene expression of *ERAP2*, but this is counterbalanced at other times by negative consequences (e.g., Crohn disease,

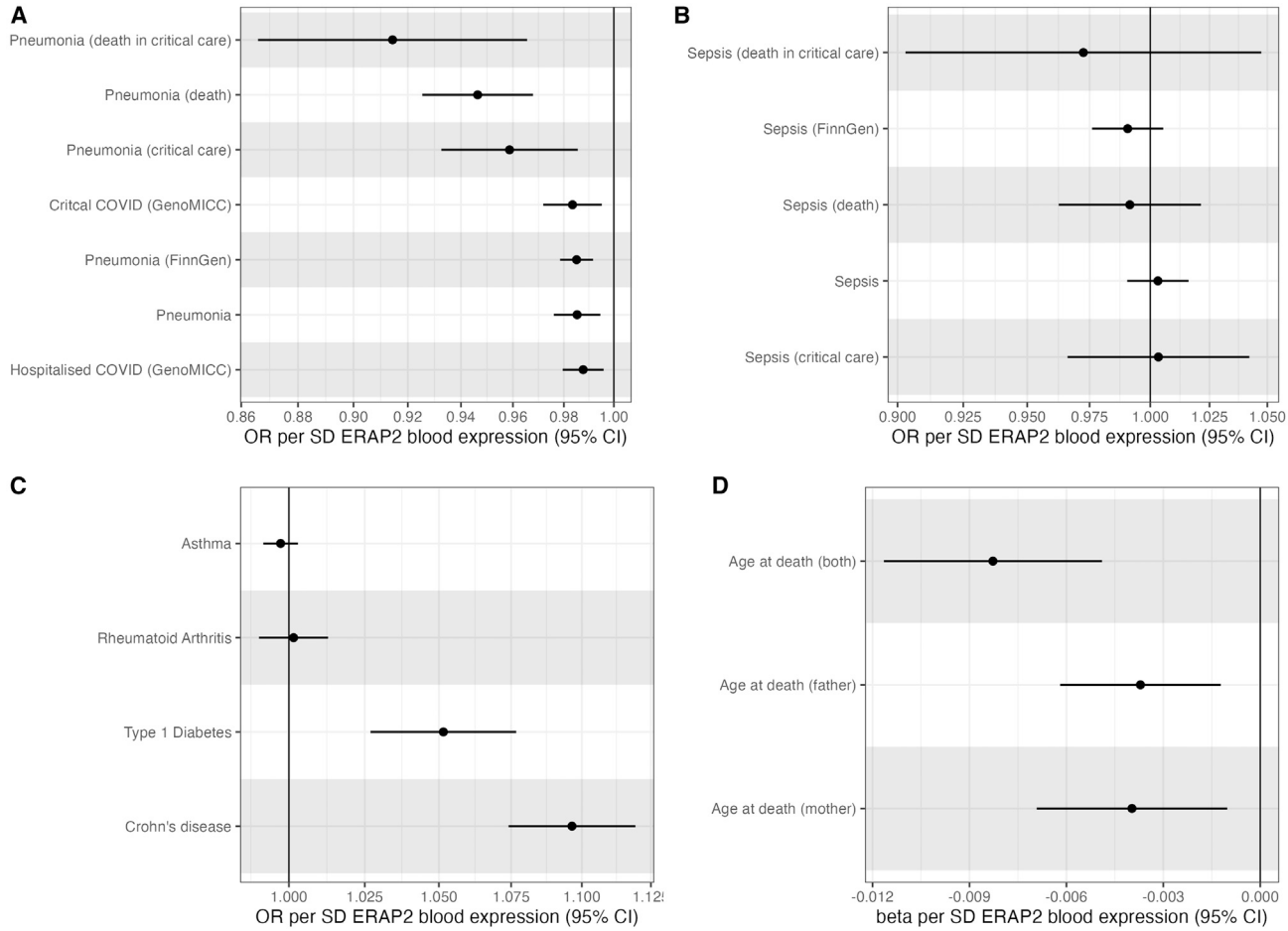

Figure 3. Inverse variance weighted MR estimates for the effect of *ERAP2* whole blood expression on outcomes
Effect estimates are on the scale of normalized gene expression. Results for (A) respiratory infection, (B) sepsis, (C) autoimmune disease, (D) parental longevity.

type 1 diabetes). Previous literature has identified that changes in *ERAP2* expression are associated with changes in survival from cancer, although one study identified increased survival[43] with increased *ERAP2* expression, and the other decreased survival,[44] while there are known genetic associations such as psoriasis[45] and juvenile idiopathic arthritis[29] at this locus. Additionally, the association with birdshot chorioretinopathy is dependent upon an interaction with HLA, further complicating analyses.[6,46] In that context, although we can speculate about the reason for balancing selection, it is difficult to provide a definitive answer, although intermittent pathogen pressure provides a plausible cause.

The major limitation of this work—in line with other genetic association studies in infection—is the challenge in defining cases of infection, which differ between and across studies. Additionally, there are assumptions and challenges of interpreting MR which are discussed extensively elsewhere,[30,20] but are more challenging at a locus such as *ERAP2* which has complex biology and a complex genetic background, with 50% of all variation in gene expression explained by a single haplotype. However, our analyses accounting for haplotype showed similar results, suggesting that our results are not simply due to haplotype. In common with all MR studies, we additionally rely on the assumptions of Mendelian randomization. In particular, we cannot confirm whether our genetic variation acts through *ERAP2* expression (the exclusion restriction assumption), nor whether the genetic variation covaries with confounders of our outcome (e.g., the included SNPs have some population structure that is coincident with a confounder that explains the association with disease). The nature of variation at this locus, with a large number of independent, common, low-frequency (minor allele frequency 0.01–0.05) variants with strong associations with *ERAP2* expression suggests an unusual genetic background and is worthy of further investigation, but should not in principle bias estimates from MR.

There are three clear implications of our study. First, these data confirm the hypotheses raised by other studies[9–12]: there is likely a balancing effect at *ERAP2* driven by protection from infection with increased risk of Crohn disease and type 1 diabetes. Secondly, therapeutics to target *ERAP1* and *ERAP2* are in development (to target Crohn disease and cancer[47]). These data suggest that targeting *ERAP2* may lead to increased rates of respiratory infection and that trials should measure and quantify infection outcomes. Finally, these findings show the potential opportunities in contemporary infection genetics: identifying and prioritizing potential gene targets using DNA from samples that are centuries old, quantifying associations in modern datasets, and identifying potential off-target effects.

Conclusion

Variation in *ERAP2* is associated with respiratory infection. In particular, the C allele of rs2549794 appears to be protective, with increasing effect estimates with increasing severity of infection. This effect was independent of ERAP2 haplotype. In supporting Mendelian randomization analyses, increased expression of *ERAP2* is protective for respiratory infection but increase the odds of some autoimmune diseases and are associated with reduced parental longevity. These results suggest ongoing balancing selection at this locus driven by autoimmune and infectious disease.

Data and code availability

All data available to replicate this analysis are publicly available. For ease, curated data are available at the authors' GitHub (https://github.com/gushamilton/ERAP2) allowing ease of replication. We do not provide code as all results can be generated by simply plotting the harmonized data or by using the TwoSampleMR package.[1]

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.02.008.

Acknowledgments

F.H.'s time was funded by the GW4-CAT Wellcome Doctoral Fellowship Scheme (222894/Z/21/Z). UK Biobank was funded by the Wellcome Trust, the Medical Research Council, the NIHR, and a variety of other charities (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/our-funding). FinnGen is a public-private partnership (https://www.finngen.fi/en/access_results) funded by multiple institutions across Finland. The authors want to acknowledge the participants and investigators of the FinnGen study. A.J.M. is a NIHR Academic Clinical Lecturer and supported by the Oxford Biomedical Research Centre (BRC). J.K.B. gratefully acknowledges funding support from a Wellcome Trust Senior Research Fellowship (223164/Z/21/Z), UKRI grants MC_PC_20004, MC_PC_19025, MC_PC_1905, MRNO2995X/1, and MC_PC_20029, Sepsis Research (Fiona Elizabeth Agnew Trust), a BBSRC Institute Strategic Program Grant to the Roslin Institute (BB/P013732/1, BB/P013759/1), and the UK Intensive Care Society. We gratefully acknowledge the support of Baillie Gifford and the Baillie Gifford Science Pandemic Hub at the University of Edinburgh. P.G.'s time was funded by the Ser Cymru programme, the Welsh Government, and the EU-ERDF. N.J.T. is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Center (BRC-1215-2001) and the MRC Integrative Epidemiology Unit (MC_UU_00011/1), and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019). T.P. was funded by the Wellcome Trust Clinical Career Development Fellowship (222098/Z/20/Z). G.D.S. works within the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the Medical Research Council (MC_UU_00011/1).

This research was funded in whole, or in part, by the Wellcome Trust (222894/Z/21/Z). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Declaration of interests

No authors declare any conflicts of interest.

Received: November 9, 2022
Accepted: February 6, 2023
Published: March 7, 2023

References

1. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., and Tan, V.Y. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife *7*, e34408. https://doi.org/10.7554/eLife.34408.
2. 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. https://doi.org/10.1016/j.humimm.2019.02.011.
3. Saulle, I., Vicentini, C., Clerici, M., and Biasin, M. (2020). An overview on ERAP roles in infectious diseases. Cells *9*, 720. https://doi.org/10.3390/cells9030720.
4. Kim, S., Lee, S., Shin, J., Kim, Y., Evnouchidou, I., Kim, D., Kim, Y.K., Kim, Y.E., Ahn, J.H., Riddell, S.R., et al. (2011). Human cytomegalovirus microRNA miR-US4-1 inhibits CD8(+) T cell responses by targeting the aminopeptidase ERAP1. Nat. Immunol. *12*, 984–991. https://doi.org/10.1038/ni.2097.
5. Lorente, E., Fontela, M.G., Barnea, E., Martín-Galiano, A.J., Mir, C., Galocha, B., Admon, A., Lauzurica, P., and López, D. (2020). Modulation of natural HLA-B*27:05 ligandome by ankylosing spondylitis-associated endoplasmic reticulum aminopeptidase 2 (ERAP2). Mol. Cell. Proteomics *19*, 994–1004. https://doi.org/10.1074/mcp.RA120.002014.
6. 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. https://doi.org/10.1093/hmg/ddy319.
7. 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. https://doi.org/10.1038/ng.2520.
8. Evans, D.M., Spencer, C.C.A., Pointon, J.J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A., Pirinen, M., Stone, M.A., et al. (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*, 761–767. https://doi.org/10.1038/ng.873.
9. 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. https://doi.org/10.1371/journal.pgen.1001157.
10. Cagliani, R., Riva, S., Biasin, M., Fumagalli, M., Pozzoli, U., Lo Caputo, S., Mazzotta, F., Piacentini, L., Bresolin, N., Clerici, M., and Sironi, M. (2010). Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. Hum. Mol. Genet. *19*, 4705–4714. https://doi.org/10.1093/hmg/ddq401.
11. 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 antiviral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. Genome Res. *28*, 1812–1825. https://doi.org/10.1101/gr.240390.118.
12. Klunk, J., Vilgalys, T.P., Demeure, C.E., Cheng, X., Shiratori, M., Madej, J., Beau, R., Elli, D., Patino, M.I., Redfern, R., and DeWitte, S.N. (2022). Evolution of immune genes is associated with the Black Death. Nature, 1–8. https://www.nature.com/articles/s41586-022-05349-x.
13. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., GTEx Consortium, and Nicolae, D.L., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098. https://doi.org/10.1038/ng.3367.
14. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986. https://doi.org/10.1038/ng.3359.
15. Zhang, S., Liu, S., Liu, N., Li, C., Wang, H., Shi, L., Zhang, X., Bao, L., Yao, Y., and Shi, L. (2020). Polymorphisms in ERAP1 and ERAP2 genes are associated with tuberculosis in the Han Chinese. Front. Genet. *11*, 566190. https://doi.org/10.3389/fgene.2020.566190.
16. Biasin, M., Sironi, M., Saulle, I., de Luca, M., la Rosa, F., Cagliani, R., Forni, D., Agliardi, C., lo Caputo, S., Mazzotta, F., et al. (2013). Endoplasmic reticulum aminopeptidase 2 haplotypes play a role in modulating susceptibility to HIV infection. AIDS *27*, 1697–1706. https://doi.org/10.1097/QAD.0b013e3283601cee.
17. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
18. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., and Pervjakova, N. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at bioRxiv. https://doi.org/10.1101/447367v1.
19. Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., et al. (2021). Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. *53*, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w.
20. Gkatzionis, A., Burgess, S., and Newcombe, P.J. (2022). Statistical methods for cis-Mendelian randomization with two-sample summary-level data. Genet. Epidemiol. *47*, 3–25. https://doi.org/10.1002/gepi.22506.
21. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209. https://doi.org/10.1038/s41586-018-0579-z.

22. Mitchell, R., Hemani, G., Dudding, T., Corbin, L., Harrison, S., and Paternoster, L. (2019). UK Biobank genetic data: MRC-IEU quality control, version 2. https://data.bris.ac.uk/data/dataset/1ovaau5sxunp2cv8rcy88688v.
23. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. Nature *591*. 591, 98. https://doi.org/10.1038/s41586-020-03065-y.
24. Forgetta, V., Manousaki, D., Istomine, R., Ross, S., Tessier, M.C., Marchand, L., Li, M., Qu, H.Q., Bradfield, J.P., Grant, S.F.A., et al. (2020). Rare genetic variants of large effect influence risk of type 1 diabetes. Diabetes *69*, 784–795. https://doi.org/10.2337/db19-0831.
25. Valette, K., Li, Z., Bon-Baret, V., Chignon, A., Bérubé, J.C., Eslami, A., Lamothe, J., Gaudreault, N., Joubert, P., Obeidat, M., et al. (2021). Prioritization of candidate causal genes for asthma in susceptibility loci derived from UK Biobank. Commun. Biol. *4*, 700. https://doi.org/10.1038/s42003-021-02227-6.
26. Ha, E., Bae, S.-C., and Kim, K. (2021). Large-scale meta-analysis across East Asian and European populations updated genetic architecture and variant-driven biology of rheumatoid arthritis, identifying 11 novel susceptibility loci. Ann. Rheum. Dis. *80*, 558–565. https://doi.org/10.1136/annrheumdis-2020-219065.
27. Pilling, L.C., Kuo, C.-L., Sicinski, K., Tamosauskaite, J., Kuchel, G.A., Harries, L.W., Herd, P., Wallace, R., Ferrucci, L., and Melzer, D. (2017). Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. Aging *9*, 2504–2520. https://doi.org/10.18632/aging.101334.
28. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. *53*, 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.
29. 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. https://doi.org/10.1038/ng.2614.
30. Sanderson, E., Glymour, M.M., Holmes, M.V., Kang, H., Morrison, J., Munafò, M.R., Palmer, T., Schooling, C.M., Wallace, C., Zhao, Q., and Davey Smith, G. (2022). Mendelian randomization. Nature Reviews Methods Primers *2*, 1–21. https://www.nature.com/articles/s43586-021-00092-5.
31. Bowden, J., Spiller, W., Del Greco M, F., Sheehan, N., Thompson, J., Minelli, C., and Davey Smith, G. (2018). Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. Int. J. Epidemiol. *47*, 1264–1278. https://doi.org/10.1093/ije/dyy101.
32. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.
33. Skrivankova, V.W., Richmond, R.C., Woolf, B.A.R., Davies, N.M., Swanson, S.A., VanderWeele, T.J., Timpson, N.J., Higgins, J.P.T., Dimou, N., Langenberg, C., et al. (2021). Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ *375*, n2233. https://doi.org/10.1136/bmj.n2233.
34. Hamilton, F., Thomas, M., Arnold, D., Palmer, T.M., Moran, E., Mentzer, A.J., Maskell, N.A., Baillie, J.K., Summers, C., Hingorani, A.D., and MacGowan, A.P. (2022). Therapeutic potential of IL6R blockade for the treatment of sepsis and sepsis-related death: Findings from a Mendelian randomisation study. Preprint at bioRxiv. https://www.medrxiv.org/content/10.1101/2022.07.14.22277638v1.abstract.
35. Kurki, M.I., Karjalainen, J., Palta, P., Sipilä, T.P., Kristiansson, K., Donner, K., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., and Loukola, A. (2022). FinnGen: Unique genetic insights from combining isolated population and national health register data. Preprint at bioRxiv. https://www.medrxiv.org/content/10.1101/2022.03.03.22271360v1.abstract.
36. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of ANthropometric Traits GIANT Consortium; and DIAbetes Genetics Replication And Meta-analysis DIAGRAM Consortium, Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375. https://doi.org/10.1038/ng.2213.
37. 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. https://doi.org/10.1038/nature15393[.
38. Tedeschi, V., Paldino, G., Paladini, F., Mattorre, B., Tuosto, L., Sorrentino, R., and Fiorillo, M.T. (2020). 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. *21*, 9608. https://doi.org/10.3390/ijms21249608.
39. Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G.R., Chiche, J.D., Coopersmith, C.M., et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA *315*, 801–810. https://doi.org/10.1001/jama.2016.0287.
40. van der Poll, T., Shankar-Hari, M., and Wiersinga, W.J. (2021). The immunology of sepsis. Immunity *54*, 2450–2464. https://doi.org/10.1016/j.immuni.2021.10.012.
41. Vitulano, C., Tedeschi, V., Paladini, F., Sorrentino, R., and Fiorillo, M.T. (2017). The interplay between HLA-B27 and ERAP1/ERAP2 aminopeptidases: from anti-viral protection to spondyloarthritis. Clin. Exp. Immunol. *190*, 281–290. https://doi.org/10.1111/cei.13020.
42. Liu, S., Cao, D., Shen, Y., Li, Y., Li, Y., Shi, L., Yu, J., Li, C., Zhang, X., Sun, M., and Yao, Y. (2017). The ERAP gene is associated with HCV chronic infection in a Chinese Han population. Hum. Immunol. *78*, 731–738. https://doi.org/10.1016/j.humimm.2017.10.005.
43. Yang, Z., Tian, H., Bie, F., Xu, J., Zhou, Z., Yang, J., Li, R., Peng, Y., Bai, G., Tian, Y., et al. (2021). ERAP2 Is Associated With Immune Infiltration and Predicts Favorable Prognosis in SqCLC. Front. Immunol. *12*, 788985. https://doi.org/10.3389/fimmu.2021.788985.
44. Lim, Y.W., Chen-Harris, H., Mayba, O., Lianoglou, S., Wuster, A., Bhangale, T., Khan, Z., Mariathasan, S., Daemen, A., Reeder, J., et al. (2018). Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration.

Proc. Natl. Acad. Sci. USA *115*, E11701–E11710. https://doi.org/10.1073/pnas.1804506115.

45. Yin, X., Low, H.Q., Wang, L., Li, Y., Ellinghaus, E., Han, J., Estivill, X., Sun, L., Zuo, X., Shen, C., et al. (2015). Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. Nat. Commun. *6*, 6916. https://doi.org/10.1038/ncomms7916.
46. Kuiper, J.J.W., 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.F., et al. (2014). A genome-wide association study identifies a functional ERAP2 haplotype associated with birdshot chorioretinopathy. Hum. Mol. Genet. *23*, 6081–6087. https://doi.org/10.1093/hmg/ddu307.
47. 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. https://doi.org/10.1016/j.ejmech.2020.113053.