A single endoplasmic reticulum aminopeptidase-1 protein allotype is a strong risk factor for Behçet's disease in HLA-B*51 carriers

Masaki Takeuchi, Michael J Ombrello, Yohei Kirino, Burak Erer, Ilknur Tugal-Tutkun, Emire Seyahi, Yilmaz Özyazgan, Norman Watts, Ahmet Gül, Daniel L Kastner, Elaine F Remmers

Online supplementary text:

Patients and controls: The individuals included in this study were either in the Turkish discovery collection or the Turkish replication collection described in the Kirino et al 2013 *Nature Genetics* report. The samples in this study (1,900 cases and 1,779 controls) are fewer than the combined Turkish discovery and Turkish replication collections because some samples were not available for genotyping or failed to pass our stringent quality control requirements. All the cases and controls were from Turkey and were genetically similar, i.e. there were no genetic outliers on the basis of principal components analysis of the genotypes of non-MHC markers genotyped on the same platform as the *ERAP1* region markers (Illumina Immunochip). All cases met the International Study Group criteria for Behçet's disease (Lancet 1990, 335:1078–1080); the controls were healthy and reported no first degree relatives with Behçet's disease. Table S1 lists the patient characteristics. All participants provided signed informed consent and the local ethics commission approved the genetic disease association analysis for the whole genome.

Legends to online supplementary figures:

Figure S1: Linkage disequilibrium in 1,779 Turkish controls among 10 non-synonymous coding ERAP1 SNPs. a) Pairwise marker D' values are in boxes connecting the markers. The numbers in the boxes are the D' value times 100. Empty boxes have D' = 1.00. b) Pairwise marker r^2 values are in boxes connecting the markers. The numbers in the boxes are the r^2 values times 100. The figures were generated with Haploview 4.2 (https://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview).

Figure S2: Behçets disease association of imputed *ERAP1* region SNPs. P-values from the results of a recessive genotype association test are shown for the imputed SNP data from 1,900 cases and 1,779 controls from Turkey. The green horizontal line shows the significance level of

homozygosity for Hap10 ($P = 3.12 \times 10^{-6}$) in 1,876 cases and 1,761 controls with high probability predicted haplotypes. Two noncoding SNPs with somewhat greater significance, but lower odds ratios are labeled. The x axis shows the positions of the SNPs on chromosome 5; transcript maps are shown below the plot.

Online Supplementary Tables

Table S1: Demographic and clinical characteristics of patients with Behçet's disease

1A. Patients from the first GWAS (1130 out of 1215)

Characteristic	*Frequency*
Male / Female, n (%)	611 / 519 (54.1 / 45.9)
Mean age ± SD (range)	39.0 ± 12.0 (13-79)
Recurrent oral aphthous ulcers (%)	100
Genital ulcers (%)	72.8
Folliculitis (%)	79.7
Erythema nodosum (%)	50.6
Pathergy reaction (n=945) (%)	80.0
Uveitis (%)	35.8
Arthritis	48.2
Vascular involvement (%)	25.3
Neurologic involvement (%)	8.6
Intestinal involvement (%)	1.1
Positive family history (%)	19.6

1B. Patients from the replication group (770 out of 838)

Characteristic	*Frequency*
Male / Female, n (%)	404 / 366 (52.5 / 47.5)
Mean age ± SD (range)	38.4 ± 11.0 (10-74)
Recurrent oral aphthous ulcers (%)	100
Genital ulcers (%)	82.9
Folliculitis (%)	73.5
Erythema nodosum (%)	57.9
Pathergy reaction (n=643) (%)	49.3
Uveitis (%)	43.6
Arthritis	39.6
Vascular involvement (%)	19.6
Neurologic involvement (%)	5.5
Intestinal involvement (%)	1.7
Positive family history (%)	18.3

1C. Whole study group (n= 1900)

Characteristic	*Frequency*
Male / Female, n (%)	1015 / 885 (53.4 / 46.6)
Mean age ± SD (range)	38.7 ± 11.6 (10-79)
Recurrent oral aphthous ulcers (%)	100
Genital ulcers (%)	77.8
Folliculitis (%)	78.2
Erythema nodosum (%)	54.2
Pathergy reaction (n=1588) (%)	67.6
Uveitis (%)	39.4
Arthritis	45.8
Vascular involvement (%)	23.2
Neurologic involvement (%)	7.4
Intestinal involvement (%)	1.3
Positive family history (%)	19.2

Table S2: Frequency of *ERAP1* non-synonymous coding SNPs with minor allele frequency greater than 0.01 in the 1000 Genomes Project EUR superpopulation and in Turkish controls.

SNP ID	*Amino acid change (ancestral/non-ancestral)**	*Nucleotide change (ancestral/non-ancestral)*	*Codon change*	*Non-ancestral allele frequency in EUR superpopulation Non-ancestral alleles/total alleles (%)*	*Non-ancestral allele frequency in Turkish controls Non-ancestral alleles/total alleles (%)*
rs72773968	Thr12Ile	C/T	T (ACC) → I (ATC)	139/1006 (13.8)	393/3558 (11.0)
rs3734016	Glu56Lys	G/A	E (GAG) → K (AAG)	41/1006 (4.1)	86/3514 (2.4)
rs26653	Pro127Arg	C/G	P (CCT) → R (CGT)	284/1006 (28.2)	1281/3558 (36.0)
rs26618	Ile276Met	A/G	I (ATA) → M (ATG)	220/1006 (21.9)	846/3558 (23.8)
rs27895	Gly346Asp	G/A	G (GGC) → D (GAC)	64/1006 (6.4)	297/3510 (8.5)
rs2287987	Met349Val	A/G	M (ATG) → V (GTG)	226/1006 (22.5)	496/3526 (14.1)
rs30187	Lys528Arg	A/G	K (AAG) → R (AGG)	654/1006 (65.0)	2202/3558 (61.9)
rs10050860	Asp575Asn	G/A	D (GAC) → N (AAC)	230/1006 (22.9)	514/3558 (14.4)
rs17482078	Arg725Gln	G/A	R (CGA) → Q (CAA)	225/1006 (22.4)	508/3558 (14.3)
rs27044	Gln730Glu	C/G	Q (CAA) → E (GAA)	719 (71.5)	2649/3532 (75.0)

*The ancestral allele is defined as the allele present in the chimpanzee sequence.

Table S3: Common *ERAP1* coding variant haplotypes (with greater than 1% frequency) and frequency-based associations of haplotypes with Behçet's disease in 3,752 case haplotypes and 3,522 control haplotypes from Turkey.

Coding haplo-type[a]	*Amino acid position*										*Hap freq cases n (%)*	*Hap freq ctrls n (%)*	*P-value*[b]	*Homozyg hap odds ratio (95% CI)*
	12	56	127	276	346	349	528	575	725	730				
Hap1	Ile[c]	Glu	Pro	Ile	Gly	Met	Lys	Asp	Arg	Gln	388 (10.3)	383 (10.9)	4.60E-01	0.95 (0.81-1.10)
Hap2	Thr	Glu	Arg	Ile	Gly	Met	Lys	Asp	Arg	Gln	458 (12.2)	491 (13.9)	2.82E-02	0.86 (0.75-0.98)
Hap3	Thr	Glu	Arg	Ile	Gly	Met	Lys	Asp	Arg	Glu	501 (13.4)	454 (12.9)	5.59E-01	1.04 (0.91-1.19)
Hap5	Thr	Glu	Arg	Ile	Asp	Met	Arg	Asp	Arg	Glu	321 (8.6)	292 (8.3)	6.85E-01	1.03 (0.88-1.22)
Hap6	Thr	Glu	Pro	Ile	Gly	Met	Arg	Asp	Arg	Glu	513 (13.7)	431 (12.2)	6.87E-02	1.14 (0.99-1.30)
Hap7	Thr	Lys	Pro	Ile	Gly	Met	Arg	Asp	Arg	Glu	106 (2.8)	85 (2.4)	2.72E-01	1.18 (0.88-1.57)
Hap8	Thr	Glu	Pro	Met	Gly	Met	Arg	Asp	Arg	Glu	790 (21.1)	840 (23.9)	4.28E-03	0.85 (0.76-0.95)
Hap10	Thr	Glu	Pro	Ile	Gly	Val	Arg	Asn	Gln	Glu	631 (16.8)	504 (14.3)	3.23E-03	1.21 (1.07-1.38)

[a]Haplotype numbers according to Ombrello et al, 2015. The 8 haplotypes listed account for 98.8% of all the haplotypes in the Turkish controls.

[b]Significance after Bonferroni correction is P < 3.12E-03 (0.05/[8 haplotypes x 2 models]). No haplotypes were significant.

[c]Yellow highlight indicates the non-ancestral amino acid; the ancestral amino acid is the amino acid found in the chimpanzee sequence.

Online Supplementary Figures

Figure S1

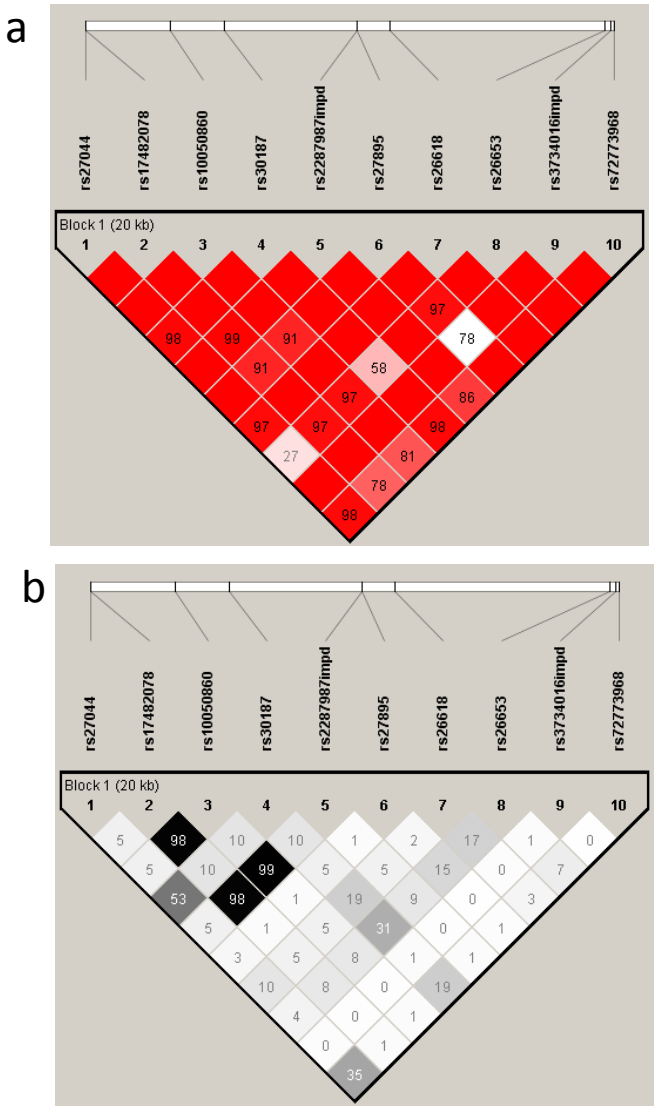
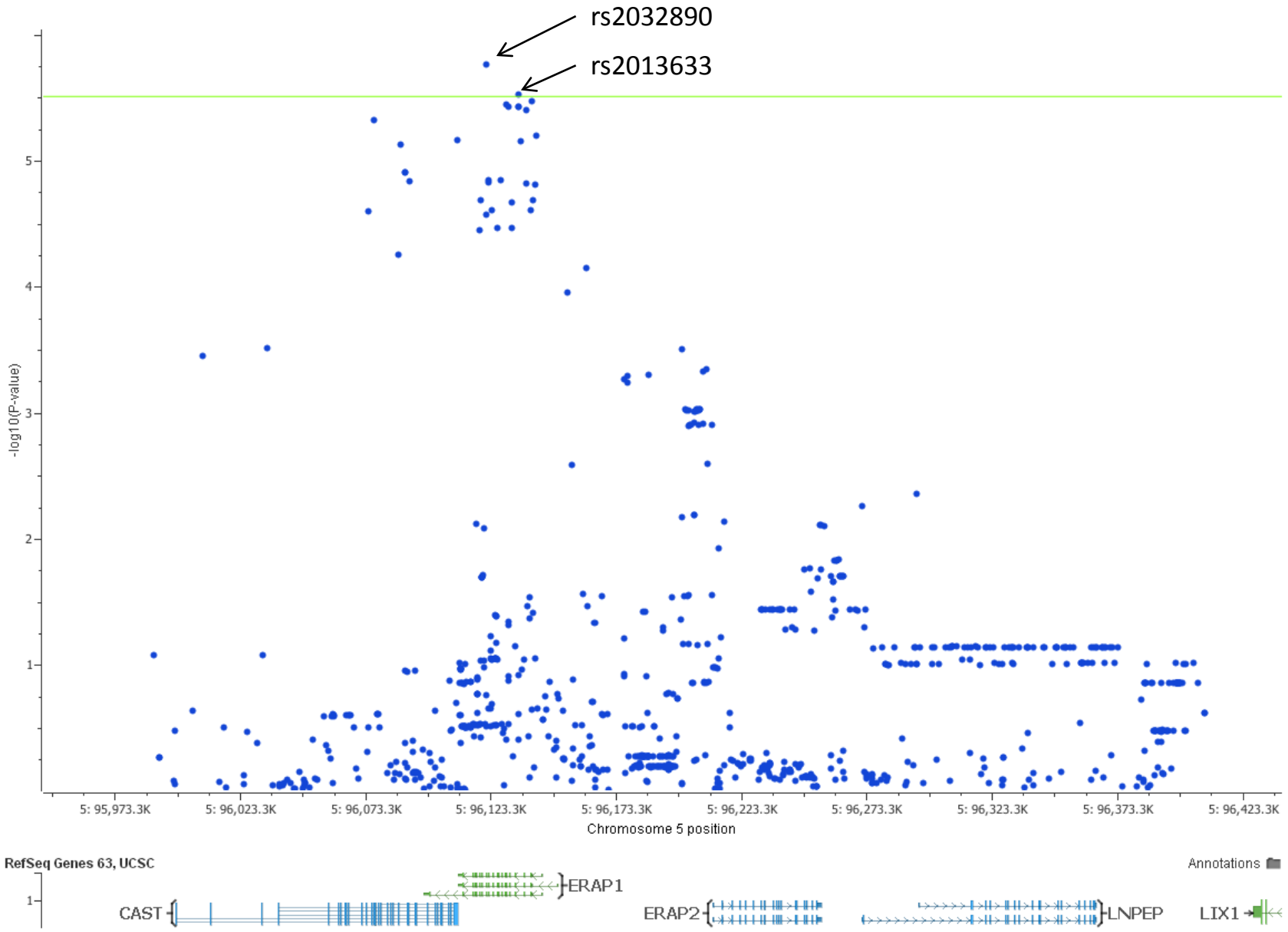


Figure S2


rs2032890
rs2013633
-log10(P-value)
5
4
3
2
1
5: 95,973.3K
5: 96,023.3K
5: 96,073.3K
5: 96,123.3K
5: 96,173.3K
5: 96,223.3K
5: 96,273.3K
5: 96,323.3K
5: 96,373.3K
5: 96,423.3K
Chromosome 5 position
RefSeq Genes 63, UCSC
Annotations
CAST
ERAP1
ERAP2
LNPEP
LIX1