Supporting Information

Ombrello et al. 10.1073/pnas.1406575111

SI Materials and Methods

Patients. Patients with Behçet disease (BD) and healthy control subjects were recruited as previously described (1). Briefly, consecutively evaluated BD patients fulfilling the International Study Group Criteria for BD (2) were enrolled, and samples were collected at the BD clinic of the Istanbul Faculty of Medicine, Division of Rheumatology. Healthy Turkish blood bank donors or healthy volunteers, selected by geographic location to match the place of birth of the patients, were recruited as healthy control subjects. Control subjects were screened to exclude individuals with any common manifestation of BD or with a family history of BD. All participants provided written informed consent, and the study was approved by the Ethics Committee of the Istanbul Faculty of Medicine.

SNP Genotyping. Genomic DNA from each subject was extracted from venous blood samples using a MagNA Pure instrument with MagNA Pure Compact DNA isolation kit (Roche), and SNP genotyping was performed subsequently using Infinium Human CNV370 bead arrays and a BeadArray Reader (Illumina) according to the manufacturer's recommendations. Markers with a call rate <0.95, discordance between the reported and chromosomal gender, or cryptic relatedness with another subject (pihat ≥0.125) were excluded, as were markers with Hardy–Weinberg equilibrium $P < 1 \times 10^{-5}$ or minor allele frequency (MAF) <0.01. The subset of 2,832 markers residing within the MHC [chromosome 6:29,500,000–33,000,000; human genome 19 (hg19)] then were extracted for subsequent analysis. Quality control and filtering operations were performed using SVS7 (Golden Helix).

Direct Typing of the HLA-B Locus. HLA-B locus typing was performed on genomic DNA using LabTYPE SSO Class I B locus assays (One Lambda) and a Bio-Plex 200 suspension array system (BioRad). For each sample, the HLA-B types were coded as biallelic markers, and associations were assessed with logistic regression in SVS7, applying a Bonferroni-corrected significance threshold accounting for 31 HLA-B types ($P < 1.6 \times 10^{-3}$). Unsupervised stepwise logistic regression with forward selection (cutoff, $P < 1.6 \times 10^{-3}$) and supervised stepwise conditional analyses of HLA-B types were performed using SVS7.

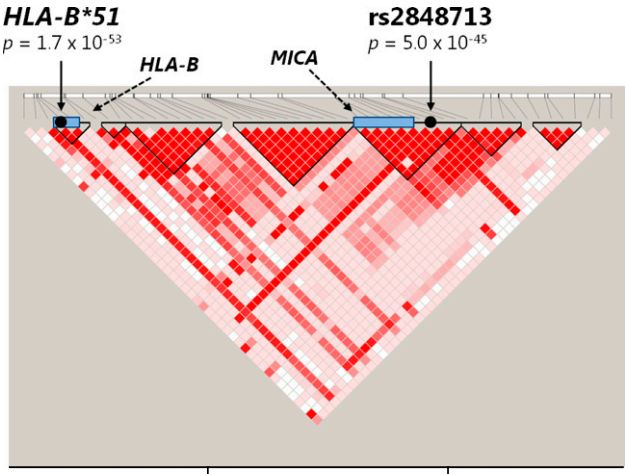
Haplotype Analysis of *HLA-B/MICA* Region. Haplotypes of SNPs from the *HLA-B*/*MICA* region with MAF >0.2 were assembled into haplotypes together with directly ascertained HLA-B alleles, and haplotype association testing was performed by χ^2 testing, as implemented in Haploview software (3).

SNP Genotype Imputation. Using the 2,832 directly ascertained MHC region SNPs as a foundation, SNP imputation was performed using ShapeIT (4) and IMPUTE2 (5) software and using the 1,000 Genomes Project phase 1 integrated dataset as the reference dataset, as previously described (6). Imputation accuracy was assessed internally by independently masking and imputing every genotyped SNP and calculating the overall concordance rate between the directly ascertained allele and the masked/imputed allele. The imputed SNP data were filtered to remove rare markers (MAF <0.01) and markers that were imputed with poor quality (IMPUTE info score <0.8). For SNPs with an info score >0.8, our masked imputation produced results that were 99.6% concordant with the directly ascertained alleles. Quality control and filtering operations were performed using SVS7 and PLINK 1.07 (7) software packages. Frequentist association testing and conditional analyses under the additive, dominant, and recessive models were performed using SNPTESTv2. Significance was determined using the genome-wide significance threshold, adjusted for three models ($P < 1.7 \times 10^{-8}$).

Imputation of Classical HLA Types and Amino Acid Residues. A set of 2,832 directly ascertained MHC SNPs also was used to impute classical HLA alleles and their corresponding amino acid sequences using SNP2HLA (8) and reference data collected by the Type I Diabetes Genetic Consortium, as described (9). This dataset included a panel of 2,537 MHC SNPs that were selected to tag the entire MHC region, together with classical types for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, and *HLA-DPB1* at four-digit resolution in 2,767 unrelated individuals of European ancestry. Additionally, the high-resolution HLA types of the reference panel were recoded into the corresponding amino acid sequences, and each residue at every polymorphic position was included in the imputation reference set as a biallelic marker. Multivariate logistic regression analysis of imputed HLA alleles was performed using PLINK-format dosage data in SVS7, and significance was determined using a Bonferroni correction for 101 imputed HLA types ($P < 5 \times 10^{-4}$). Haplotypes of two-digit MHC class I alleles were generated using the expectation maximization algorithm, as implemented in SVS7, and haplotype association testing was performed using logistic regression in SVS7.

Analysis of HLA-B and HLA-A Amino Acid Positions. To identify the effects of individual amino acid positions, we performed supervised stepwise logistic regression of probabilistic doses of polymorphic amino acid positions of HLA-B and HLA-A using SVS7. We encoded each polymorphic position as a series of binary markers, but because the allelic markers at any given position were not independent of one another, it was necessary to use an omnibus strategy in our regression to control for the effects of all alleles at a given position simultaneously. Therefore, at each step of our analysis, we included all the allelic markers at a given position as covariates in the ensuing step of the regression analysis. Modeling of HLA-A and HLA-B was performed using PyMol software (DeLano Scientific) and previously defined crystal structures of HLA-B*51:01 (10) and HLA-A*03:01 (11).

1. Remmers EF, et al. (2010) Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. *Nat Genet* 42(8):698–702.
2. International Study Group for Behçet's Disease (1990) Criteria for diagnosis of Behçet's disease. *Lancet* 335(8697):1078–1080.
3. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.
4. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181.
5. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39(7):906–913.
6. Howie B, Marchini J, Stephens M (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1(6):457–70.
7. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
8. Jia X, et al. (2013) Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* 8(6):e64683.
9. Raychaudhuri S, et al. (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 44(3):291–296.
10. Maenaka K, et al. (2000) Nonstandard peptide binding revealed by crystal structures of HLA-B*5101 complexed with HIV immunodominant epitopes. *J Immunol* 165(6):3260–3267.
11. McMahon RM, et al. (2011) Structure of HLA-A*0301 in complex with a peptide of proteolipid protein: Insights into the role of HLA-A alleles in susceptibility to multiple sclerosis. *Acta Crystallogr D Biol Crystallogr* 67(Pt 5):447–454.



Extended haplotype	OR (95%CI)	*p*	Case *f*	Control *f*
*HLA-B*51* positive	2.81 (2.44, 3.25)	9.4E-47	0.321	0.144
*HLA-B*51* negative	0.96 (0.72, 1.27)	0.764	0.042	0.044

Fig. S1. *HLA-B*51* is an essential element of the BD-associated *HLA-B*/*MICA* extended haplotype. Haplotype analysis of *HLA-B*51* together with *HLA-B*/*MICA* region SNPs with MAF >0.2 identified one haplotype comprising *HLA-B*51* and 48 SNPs that completely included both *HLA-B* and *MICA* and was strongly associated with BD. The identical 48-SNP haplotype also was observed among *HLA-B*51*–negative individuals, in whom it conferred no risk of BD.

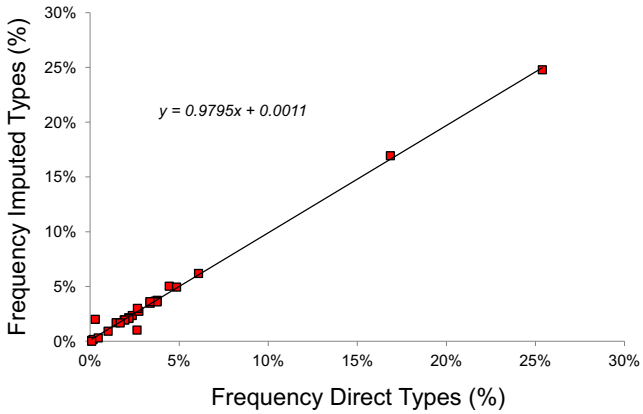


Fig. S2. Correlation of allelic frequencies of directly ascertained and imputed two-digit *HLA-B* locus types. The locations of red boxes, each of which represents a two-digit *HLA-B* antigen, demonstrate the relationship between directly typed and imputed HLA-B allotypic frequencies. The black line represents the results of linear regression which identified strong correlation between direct and imputed HLA-B types (R^2 = 0.9913).

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)
Table S5 (DOCX)
Table S6 (DOCX)
Table S7 (DOCX)