Behçet disease-associated MHC class I residues implicate antigen binding and regulation of cell-mediated cytotoxicity

Michael J. Ombrello[a,b,1], Yohei Kirino[b,c], Paul I. W. de Bakker[d,e,f,g], Ahmet Gül[h], Daniel L. Kastner[b,1], and Elaine F. Remmers[b]

[a]Translational Genetics and Genomics Unit, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892; [b]Inflammatory Disease Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; [c]Department of Internal Medicine and Clinical Immunology, Yokohama City University Graduate School of Medicine and Faculty of Medicine, Yokohama City 236-0004, Japan; [d]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; [e]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02141; Departments of [f]Medical Genetics and [g]Epidemiology, University Medical Center Utrecht, 3584 CX, Utrecht, The Netherlands; and [h]Division of Rheumatology, Department of Internal Medicine, Istanbul Faculty of Medicine, Istanbul University, Istanbul 34093, Turkey

The HLA protein, HLA-B*51, encoded by *HLA-B* in MHC, is the strongest known genetic risk factor for Behçet disease (BD). Associations between BD and other factors within the MHC have been reported also, although strong regional linkage disequilibrium complicates their confident disentanglement from *HLA-B*51*. In the current study, we examined a combination of directly obtained and imputed MHC-region SNPs, directly obtained *HLA-B* locus types, and imputed classical HLA types with their corresponding polymorphic amino acid residues for association with BD in 1,190 cases and 1,257 controls. SNP mapping with logistic regression of the MHC identified the *HLA-B/MICA* region and the region between *HLA-F* and *HLA-A* as independently associated with BD ($P < 1.7 \times 10^{-8}$). *HLA-B*51, -*A*03, -*B*15, -*B*27, -*B*49, -*B*57*, and -*A*26* each contributed independently to BD risk. We directly examined rs116799036, a noncoding SNP upstream of *HLA-B* that was recently suggested to underlie the association of *HLA-B*51* with BD, but we were unable to replicate that finding in our collection. Instead, we mapped the BD association to seven MHC class I (MHC-I) amino acid residues, including anchor residues that critically define the selection and binding of peptides to MHC-I molecules, residues known to influence MHC-I–killer immunoglobulin-like receptor interactions, and a residue located in the signal peptide of HLA-B. The locations of these variants collectively implicate MHC-I peptide binding in the pathophysiology of BD. Furthermore, several lines of evidence suggest a role for altered regulation of cellular cytotoxicity in BD pathogenesis.

HLA imputation | autoinflammation | antigen presentation | killer immunoglobulin-like receptors | natural killer cells

Behçet disease (BD) is a multisystem inflammatory disease of complex inheritance with a clinical course marked by recurrent episodes of oral and genital ulceration, severe ocular inflammation often leading to visual impairment or blindness, and a range of inflammatory lesions of the skin and the gastrointestinal, neurologic, and circulatory systems (1). The predominant BD susceptibility locus is the MHC on chromosome 6 (2, 3), which contains the strongest known risk factor for BD, the MHC class I (MHC-I) allele *HLA-B*51* (2–5). Several recent studies have expanded the list of genes or loci implicated in the pathophysiology of BD, which now includes *HLA-B*, *IL10*, *IL23R*, *HLA-A*, *CCR1*, *STAT4*, endoplasmic reticulum amino peptidase 1 (*ERAP1*), the killer lectin-like receptor cluster on chromosome 12, and, most recently, *TLR4* and *MEFV* (2, 3, 6, 7). Although these genetic studies of BD have provided new clues and insights into the pathogenesis of BD, none has provided a thorough accounting of the individual risk factors within the MHC. The lack of such a study likely reflects the absence of a BD study population of adequate size to overcome the strong linkage disequilibrium (LD) and to disentangle from *HLA-B*51* the additional risk factors within the MHC.

Multiple lines of evidence suggest that sources of BD risk, in addition to *HLA-B*51*, exist within the MHC. This evidence begins in the *HLA-B* locus, where associations between BD and several alleles in addition to *HLA-B*51* have been reported (8–10). It also has been argued that variants in or around MHC class I polypeptide-related sequence A (*MICA*), the centromeric neighbor of *HLA-B* that encodes the MHC-I chain-related sequence A, contribute to BD susceptibility (11). However, efforts to parse the effects of *MICA* and *HLA-B* alleles definitively have been confounded by their particularly strong LD (11–14). Additionally, *HLA-A* has been identified as a BD susceptibility locus in numerous studies (2, 3, 14–17), and it has been suggested that *HLA-C* contributes to BD risk, as well (14).

To understand better the sources of BD risk within the MHC, we have analyzed directly ascertained and imputed SNP genotypes, together with HLA type and amino acid data from a very large and meticulously assembled collection of Turkish subjects with BD and geographically matched, healthy Turkish individuals. Using stepwise and multivariate logistic regression, conditional analysis, and haplotype analysis, we sought to characterize the range of genetic risk factors for BD contained within the MHC.

Significance

To elucidate their role in Behçet disease (BD), we used imputation to analyze MHC class I (MHC-I) alleles and amino acids in a large case-control collection. Several MHC-I alleles influenced BD susceptibility, some conferring risk and others affording protection. The association of MHC-I with BD was mapped to six positions around the peptide-binding groove and to one position in the signal peptide. These residues are critical in defining the peptide-binding specificity of MHC-I molecules, and they also affect the engagement of MHC-I molecules by killer immunoglobulin-like receptors on natural killer and T cells. These data implicate peptide binding by MHC-I in BD pathogenesis and suggest that altered regulation of cytotoxic cells by MHC-I may be pathogenic in BD.

GENETICS

Table 1. Additive model association testing and stepwise conditional analysis of directly ascertained HLA-B antigens in BD

Covariates	Risk allele	P value[†]	OR (95% CI)
Full collection			
None	HLA-B*51	1.3×10^{-55}	3.0 (2.6, 3.4)
HLA-B*51	HLA-B*15	1.0×10^{-5}	1.9 (1.4, 2.5)
HLA-B*51 + HLA-B*15	HLA-B*27	1.0×10^{-3}	1.7 (1.2, 2.3)
HLA-B*51 + HLA-B*15 + HLA-B*27	HLA-B*49	7.5×10^{-3}	0.6 (0.4, 0.9)
	HLA-B*57	9.4×10^{-3}	1.7 (1.1, 2.6)
HLA-B*51-negative subset			
None	HLA-B*15	3.4×10^{-5}	2.0 (1.4, 2.7)
HLA-B*15	HLA-B*49	1.1×10^{-4}	0.4 (0.2, 0.7)
HLA-B*15	HLA-B*57	5.5×10^{-3}	2.0 (1.2, 3.2)
+ HLA-B*49	HLA-B*27	6.5×10^{-3}	0.4 (0.2, 0.7)

[†]After correcting for 31 directly ascertained HLA-B types, significance was defined as $P < 1.6 \times 10^{-3}$.

Results

Multiple *HLA-B* Alleles Independently Influence Susceptibility to BD. We performed association testing of directly ascertained two-digit *HLA-B* locus types in 1,190 BD cases and 1,257 healthy subjects from Turkey. Our results affirmed that *HLA-B*51* is the largest single risk factor for BD, acting in an additive fashion with an odds ratio (OR) of 3.0 per copy ($P = 1.3 \times 10^{-55}$) (Table 1 and Table S1). In addition to *HLA-B*51*, both the supervised conditional analysis (Table 1) and the unsupervised stepwise forward logistic regression analysis (Table S2) of *HLA-B* alleles in the full collection identified significant, independent effects of *HLA-B*15* and *HLA-B*27* on the risk of BD. *HLA-B*15* also was significantly associated with BD in the *HLA-B*51*–negative subset of the collection, which included 487 cases and 889 controls ($P = 3.4 \times 10^{-5}$) (Table 1). Conditional analysis accounting for the effect of *HLA-B*15* on BD risk in the *HLA-B*51*–negative subset unveiled a significant, protective effect of *HLA-B*49* on BD ($p_{regressor} = 1.1 \times 10^{-4}$) (Table 1), and, as is consistent with these findings, the model of BD risk generated by forward stepwise logistic regression included *HLA-B*15* and *HLA-B*49* (Table S2). Additionally, *HLA-B*57* trended toward

association with BD in analyses of the full collection and of the *HLA-B*51*–negative subset of the collection (Table 1).

HLA-B*51 Is the Primary BD Risk Factor in the HLA-B/MICA Region. Association testing of 32,689 directly acquired and imputed SNPs from the MHC region revealed that the strongest peak of association included *HLA-B* and *MICA* and consisted of more than 775 BD-associated SNPs whose associations exceeded the threshold for genome-wide significance ($P < 1.7 \times 10^{-8}$), with the most significant associations observed under the additive model (Fig. 1). Among these was the most strongly BD-associated SNP in the study, rs79556279 [$p_{additive} = 2.2 \times 10^{-50}$, OR 2.7 (95% confidence interval, CI, 2.3, 3.1)], which was located 4.9 kb 5′ of *HLA-B*. After controlling for the effect of rs79556279, we found that no other SNP in the *HLA-B/MICA* region was significantly associated with BD (Fig. 1A). Association testing of MHC-region SNPs conditioned on the effect of *HLA-B*51* similarly identified no significant residual association in the *HLA-B/MICA* region (Fig. 1B), and, moreover, rs79556279 was in strong LD with *HLA-B*51* [expectation–maximization r^2 (r^2_{EM}) = 0.92; expectation-maximization pairwise linkage disequilibrium (D'_{EM}) = 0.96], indicating that the effect of *HLA-B*51* underlies the observed effect of rs79556279. In fact, the majority of BD-associated markers in the *HLA-B/MICA* region were in moderate to strong LD with *HLA-B*51* (Fig. 1C).

To examine the role of rs116799036, a noncoding SNP recently proposed to underlie the effect of *HLA-B*51* on BD risk (14), we directly genotyped this marker in our collection. The association of rs116799036 with BD was seven orders of magnitude weaker than that of *HLA-B*51* (Table S3). Furthermore, in contrast to a previous report (14), the association of *HLA-B*51* with BD remained significant even after conditioning for the effect of rs116799036 ($p_{regressor} = 9.2 \times 10^{-8}$; Table S3).

In an effort to localize further the source(s) of BD risk within the *HLA-B/MICA* region, we analyzed haplotypes of *HLA-B/MICA* region SNPs [minor allele frequency (MAF) >0.2] with *HLA-B*51* using Haploview (18). We identified one haplotype, consisting of 48 *HLA-B/MICA* region SNPs and *HLA-B*51*, that was strongly associated with BD [$P = 9.4 \times 10^{-47}$; OR 2.81 (95% CI 2.44, 3.25)] (Fig. S1). Interestingly, a nearly identical form of this haplotype that bore the same 48 SNPs but lacked *HLA-B*51* was present at equal frequencies in BD patients and healthy controls, thereby conferring no demonstrable effect on BD risk ($P = 0.764$).
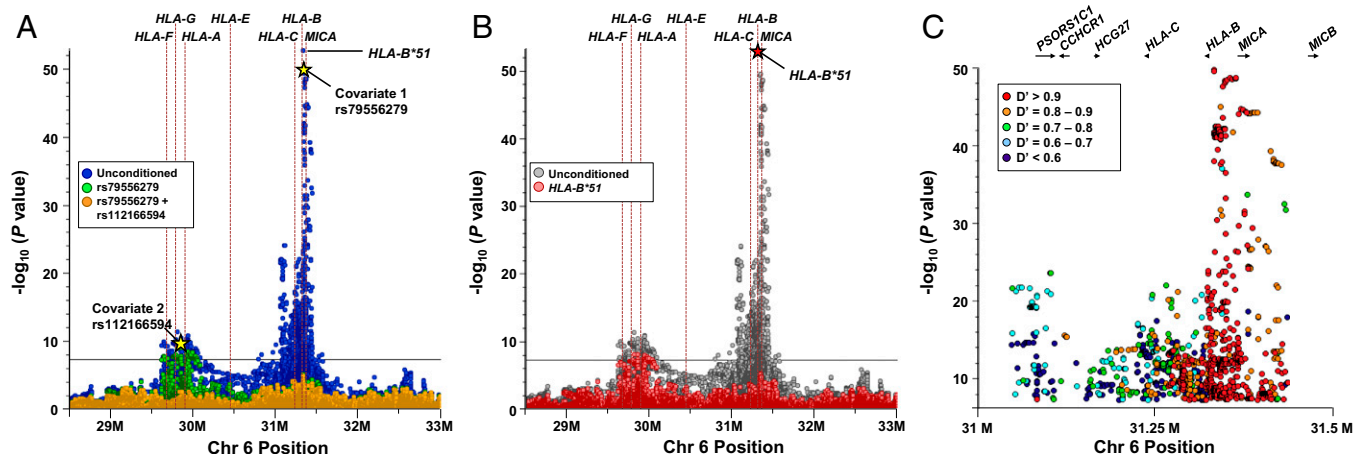
Fig. 1. *HLA-B*51* is the predominant risk allele, but variants between *HLA-F* and *HLA-A* are independently associated with BD. (*A* and *B*) The results of association testing and stepwise conditional analysis of imputed MHC region SNPs in 1,190 BD cases and 1,257 healthy control subjects from Turkey are displayed in *A*. Conditional analysis accounting for the effect of *HLA-B*51* (*B*, red dots) produced a pattern of residual association virtually identical to that seen after conditioning for rs79556279 (*A*, green dots). (*C*) Association testing results of BD-associated SNPs in proximity to *HLA-B/MICA* were plotted, and data points were color-coded to demonstrate D′ of each SNP with *HLA-B*51*.

The Region Between *HLA-F* and *HLA-A* Is an *HLA-B*51*-Independent BD Susceptibility Locus. In addition to the *HLA-B/MICA* region, association testing of MHC-region SNPs under the additive model identified two other BD-associated regions, one located telomeric to *HLA-C* and a second that included *HLA-F*, *-G*, *-H*, and *-A* (Fig. 1). Conditional analysis correcting for the effect of rs79556279 revealed that variants in the *HLA-F/HLA-A* region were independently associated with BD, whereas those in the region telomeric to *HLA-C* were not (Fig. 1 *A* and *B*). After controlling for the effect of rs79556279, the *HLA-F/HLA-A* region included 113 SNPs whose associations with BD exceeded the genome-wide significance threshold, the strongest of which was rs112166594 [$p_{additive} = 8.3 \times 10^{-10}$, OR 0.56 (95% CI 0.48, 0.67)]. The three regional SNPs most strongly associated with BD were located nearest to *HLA-H*, *HLA-G*, and *HLA-A*, respectively, whereas the majority of BD-associated SNPs in this region were located nearest to *HLA-A*. Finally, conditional analysis accounting for the effects of both rs79556279 and rs112166594 failed to identify any SNP associations whose significance met even a nominally significant threshold corrected for the 2,847 directly typed SNPs ($P < 1.7 \times 10^{-5}$) (Fig. 1*A*).

Multiple *HLA-B* and *HLA-A* Alleles Independently Influence the Risk of Developing BD. Directly ascertained MHC SNPs were used to infer HLA types at the classical MHC-I and MHC-II loci using a large reference panel of European ancestry, as previously described (19). The concordance rate between imputed and directly ascertained two-digit *HLA-B*51* alleles was 98.6%, the overall concordance rate among two-digit *HLA-B* alleles was 95.9%, and their allelic frequencies were highly correlated ($r^2 = 0.9913$) (Fig. S2). Univariate logistic regression of imputed HLA-type dosage data affirmed that *HLA-B*51* was the strongest genetic risk of BD [$P = 3.4 \times 10^{-58}$, OR 3.3 (95% CI 2.8, 3.8)] (Table 2 and Table S1) with metrics of association nearly identical to those of imputed *HLA-B*51:01* [$P = 5.1 \times 10^{-58}$, OR 3.3 (95% CI 2.8, 3.8)] (Tables S1 and S4) and very similar to those of directly typed *HLA-B*51* [$P = 1.3 \times 10^{-55}$, OR 3.0 (95% CI 2.6, 3.4)] (Table 1 and Table S1). To identify additional MHC-I alleles that influence BD risk, we performed conditional analysis and stepwise forward logistic regression of imputed HLA types in the full collection and in the *HLA-B*51*–negative subset (Table 2). After controlling for the effect of imputed *HLA-B*51* in the full collection, *HLA-A*03* was significantly protective against BD (Table 2 and Table S2). We found that *HLA-A*03* was in strong LD with the SNP most strongly BD-associated in the *HLA-A* region, rs112166594 ($r^2_{EM} = 0.99$, D′$_{EM} = 1.0$). Additionally, *HLA-B*15*, *-B*49*, *-A*26*, and *-B*27* each trended toward association with BD after correcting for the effects of *HLA-B*51* and *-A*03*, although without reaching statistical significance

(Table 2). In the *HLA-B*51*–negative subset of the population, stepwise forward logistic regression of imputed classical MHC alleles produced a model of BD risk that included *HLA-A*03* and *-B*49* (Table S2). In addition to *HLA-A*03* and *HLA-B*49*, stepwise conditional analysis also demonstrated a trend toward association between BD and *HLA-B*15* and *HLA-A*26* (Table 2). Analogous analyses of imputed four-digit MHC alleles found that, upon controlling for the effect of imputed *HLA-B*51:01* in the full study population, *HLA-A*03:01* was significantly protective against BD [$P = 2.4 \times 10^{-6}$, OR 0.6 (95% CI 0.4, 0.7)] (Table S2). Further, in the *HLA-B*51*–negative population, *HLA-B*49:01* was significantly protective against BD [$P = 1.2 \times 10^{-5}$, OR 0.6 (95% CI 0.4, 0.8)] (Table S2).

Haplotypes of imputed two-digit MHC-I alleles with frequencies >0.01 were assembled with the EM algorithm, as implemented in SVS7 (Table S5). With one exception, all *HLA-B*51*–containing haplotypes conferred risk of BD, and all *HLA-A*03*–containing haplotypes were protective against BD (Table S6). The sole exception to this rule was the haplotype bearing both *HLA-A*03* and *HLA-B*51*, the only haplotype bearing either allele that did not influence BD risk, illustrating that the protective effect of *HLA-A*03* and the risk effect of *HLA-B*51* are independent and able to counteract one another (Table S6).

Five Amino Acid Residues of HLA-B and Two Residues of HLA-A Influence BD Risk. In an effort to identify a motif that may explain the variety of protective and risk effects differentially conferred by *HLA-B* and *HLA-A* alleles in BD, polymorphic amino acid residues were derived from the imputed MHC type data, as previously described (19). Stepwise conditional analysis of the polymorphic amino acid positions of HLA-B in the full study population revealed that positions 97, 116, 152, and 67 each significantly and independently influenced the risk of developing BD (Table 3 and Fig. 2). Similar analyses of HLA-A amino acid positions identified significant and independent associations between residues 161 and 97 of HLA-A and BD risk (Table 3 and Fig. 2). When amino acids from HLA-B and HLA-A were evaluated together by stepwise conditional analysis, a model of BD risk was generated that included five of these six positions as independent influences on BD risk (Table 3). In further conditional analyses of the polymorphic amino acid positions of HLA-B and/or HLA-A in the *HLA-B*51*–negative subset (Table 3), residue −10 of HLA-B and residue 161 of HLA-A were significantly and independently associated with BD, and residue 149 of HLA-A also trended toward association with BD.

Residues 97, 116, 152, and 67 of the HLA-B protein are each located in the MHC-I antigen-binding groove (Fig. 3), where they act to define the sizes and shapes of individual peptide residues accommodated by specific MHC-I molecules (20, 21). Collectively these four MHC-I residues physically contact six of nine peptide residues (21), and residues 67 and 116 are critical anchor residues of particular importance in defining the peptide specificity of MHC-I antigen-binding grooves through their interactions with peptide positions P2 and P9, respectively (20, 21). Positions 67 and 116 also are critical determinants of the interactions of HLA-B molecules with the killer immunoglobulin-like receptors (KIR) KIR3DL1 and KIR3DS1, which regulate the activation of natural killer (NK) cells and CD8+ cytotoxic T lymphocytes (CTLs) (22). In addition, residue 67 is one of two residues at which HLA-B*51 differs from the nearly identical HLA-B protein HLA-B*52, which confers no significant effect on BD risk [$P = 0.19$, OR 0.77 (95% CI 0.5, 1.1)], further emphasizing the importance of this position. One BD-associated residue was located distant from the antigen-binding groove at position −10 of the HLA-B signal peptide. Interestingly, the HLA-B signal peptide also regulates CTL and NK cell activation, although it does so by a mechanism independent of peptide binding and KIR engagement by HLA-B (23).

MHC-I molecules bearing the Bw4 epitope of the α1-binding pocket, defined by leucine and arginine at positions 82 and 83,

Table 2. Additive model association testing and stepwise conditional analysis of imputed two-digit classic HLA alleles in BD

Covariates	Risk allele	P value[†]	OR (95% CI)
Full collection			
None	HLA-B*51	3.4×10^{-58}	3.3 (2.8, 3.8)
HLA-B*51	HLA-A*03	4.0×10^{-8}	0.6 (0.5, 0.7)
HLA-B*51 + HLA-A*03	HLA-B*15	8.7×10^{-4}	1.6 (1.2, 2.1)
HLA-B*51	HLA-B*49	2.3×10^{-3}	0.6 (0.4, 0.8)
+ HLA-A*03	HLA-A*26	3.5×10^{-3}	1.5 (1.1, 2.0)
+ HLA-B*15	HLA-B*27	5.8×10^{-3}	1.6 (1.1, 2.2)
HLA-B*51–negative subset			
None	HLA-B*49	1.1×10^{-5}	0.3 (0.2, 0.6)
HLA-B*49	HLA-A*03	1.1×10^{-4}	0.6 (0.5, 0.8)
HLA-B*49	HLA-B*15	4.5×10^{-3}	1.6 (1.1, 2.1)
+ HLA-A*03	HLA-A*26	6.7×10^{-3}	1.6 (1.1, 2.2)

[†]After correcting for 101 imputed two-digit classical HLA alleles, significance was defined as $P < 5.0 \times 10^{-4}$.

Ombrello et al.

Table 3. Stepwise conditional analysis of HLA-B and HLA-A amino acid positions in BD

HLA locus	Covariates	Effect amino acids	$P_{regressor}$	OR (95% CI)
Full collection*				
HLA-B only[†]	None	Thr 97	1.5×10^{-38}	2.3 (2.0, 2.7)
	97	Leu 116	2.5×10^{-10}	0.5 (0.4, 0.6)
	97, 116	Glu 152	6.2×10^{-8}	1.6 (1.3, 1.9)
	97, 116, 152	Phe 67	2.0×10^{-4}	1.4 (1.2, 1.6)
HLA-A only[‡]	None	Glu 161	2.0×10^{-10}	0.5 (0.4, 0.7)
	161	Arg 97	1.3×10^{-4}	1.3 (1.1, 1.4)
HLA-B and -A[§]	None	Thr 97 (B)	1.5×10^{-38}	2.3 (2.0, 2.7)
	B97	Leu 116 (B)	2.5×10^{-10}	0.5 (0.4, 0.6)
	B97, B116	Glu 161 (A)	3.0×10^{-8}	0.6 (0.5, 0.7)
	B97, B116 + A161	Glu 152 (B)	3.2×10^{-8}	1.6 (1.4, 1.9)
	B97, B116, A161 + B152	Phe 67 (B)	2.2×10^{-4}	1.4 (1.2, 1.6)
HLA-B*51–negative subset*				
HLA-B only[†]	None	Ala −10	3.3×10^{-5}	0.7 (0.6, 0.8)
HLA-A only[‡]	None	Glu 161	6.1×10^{-5}	0.6 (0.4, 0.8)
	161	Glu 149	7.8×10^{-3}	1.5 (1.1, 1.9)
HLA-B and -A[§]	None	Ala −10 (B)	3.3×10^{-5}	0.7 (0.6, 0.8)
	B-10	Glu 161 (A)	2.5×10^{-10}	0.5 (0.4, 0.6)
	B-10 + A161	Glu 149 (A)	1.5×10^{-3}	1.6 (1.2, 2.1)

*The inclusion of Bw4 as a potential covariate did not alter the results.
[†]Cutoff P value adjusted for 69 HLA-B positions = 7.2×10^{-4}.
[‡]Cutoff P value adjusted for 68 HLA-A positions = 7.4×10^{-4}.
[§]Cutoff P value adjusted for 137 HLA-A and HLA-B positions = 3.6×10^{-4}.

bind to a specific subset of KIRs, sometimes in an antigen-dependent fashion (24, 25). To investigate a potential role for Bw4 in BD, Bw4$^+$ alleles of HLA-B and HLA-A were examined. Bw4 epitopes in HLA-B molecules were a weak risk factor for BD in univariate analysis of the full collection, but Bw4 epitopes present in HLA-A molecules conferred no risk of BD (Table S7). Upon conditioning for the effect of *HLA-B*51* in the full collection, no Bw4 epitope was associated with BD; moreover, Bw4 had no effect on BD risk in the *HLA-B*51*–negative subset of the population (Table S7). Most importantly, addition of Bw4 variables to our stepwise analysis of MHC-I amino acid positions showed that Bw4 was not an independent risk factor for BD (see footnotes of Table 3).

In all, multivariate logistic regression analyses of classical MHC amino acid data identified seven MHC-I amino acid positions as risk factors for BD, implicating peptide binding by both HLA-B and HLA-A and the signal peptide of HLA-B in the pathophysiology of BD.

Discussion

By applying contemporary techniques to the examination of a large, well-matched collection of BD cases and healthy controls, this study has provided the most extensive interrogation of the HLA locus in BD to date. Our data have brought clarity to questions surrounding the relationship of *HLA-B*51* with BD while also bringing to light additional MHC-I alleles that influence BD

susceptibility. Moreover, this study has, for the first time to our knowledge, mapped the association of MHC-I molecules with BD to specific HLA-B and HLA-A amino acid residues, most of which cluster in the antigen-binding groove at positions known to influence strongly the selection and binding of peptide antigens by MHC-I molecules. These findings have important implications for the pathophysiology of BD, because peptide binding by MHC-I directly affects (*i*) the folding and stability of peptide–MHC-I complexes, which in turn determine their surface expression vs. retention in the endoplasmic reticulum; (*ii*) the recognition of peptide–MHC-I complexes by antigen-specific T-cell receptors on CTLs; and (*iii*) the engagement of KIRs on CTLs and NK cells by peptide–MHC-I complexes (26–30).

The relationship between *HLA-B*51* and BD was first identified four decades ago; since then this association has been replicated in nearly every genetic study of BD. We have affirmed this relationship in our study of a very large collection of Turkish BD patients, defining the estimated effect size as OR 3.0 (95% CI 2.6, 3.4) per allele in an additive model. *HLA-B*51* is in strong LD with genetic variants spanning the region that includes *HLA-B* and *MICA* (Fig. 1C), many of which are associated with BD. Using haplotype analysis of *HLA-B*51* and SNPs in the *HLA-B/MICA* region, we demonstrated the importance of *HLA-B*51* itself in BD risk. This observation is contrary to a recent report by Hughes et al. (14) that suggested that, instead of *HLA-B*51*, rs116799036, a noncoding variant between
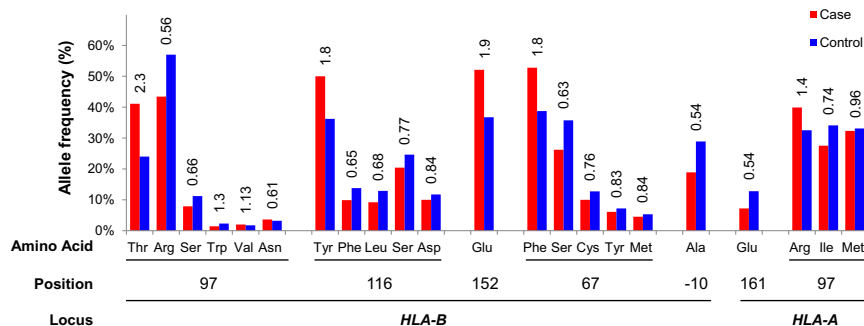
Fig. 2. Individual amino acids within HLA-A and HLA-B proteins influence the risk of BD. The allele frequencies of amino acid positions 97, 116, 152, 67, and −10 in HLA-B and positions 161 and 97 in HLA-A are plotted for cases (red) and controls (blue), and univariate ORs are displayed above the bars.
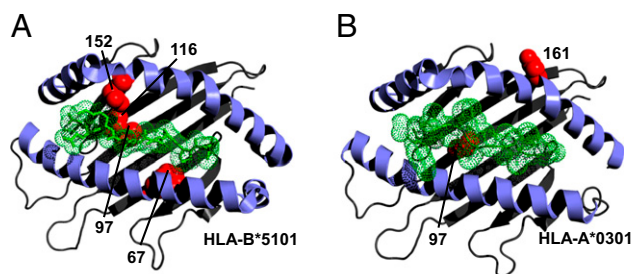
Fig. 3. BD-associated positions within MHC-I molecules cluster around the antigen-binding groove. 3D modeling of HLA-B*5101 (*A*) and HLA-A*0301 (*B*) demonstrates the clustering of BD-associated amino acid positions (red balls) in and around the antigen-binding grooves of HLA-B and HLA-A. The peptide backbone and side chains of the epitope are displayed in green. This figure was prepared with PyMol, using Protein Data Bank entries 2XPG (40) and 1E27 (41).

HLA-B and *MICA*, was the true source of BD risk in this region. In our study, *HLA-B*51* was much more strongly associated with BD than was any SNP, including rs116799036 (Fig. 1*B* and Table S3); moreover, *HLA-B*51* continued to confer significant risk of BD, even after controlling for the effect of rs116799036 (Table S3).

In comparing the results of our study with that of Hughes et al. (14), there are several important considerations. First, our study examined nearly twice as many BD cases as did the study by Hughes et al. and thus has greater statistical power to assess BD risk factors. Second, we assembled our study population prospectively with pairwise geographic matching of controls to cases to minimize population stratification (2). Third, we used a proven HLA imputation methodology that leverages a large reference panel of mixed European ancestry (19, 31), whereas Hughes et al. used a different algorithm that uses a reference panel of Northern European ancestry (32). Therefore, it is possible that differences in the imputation methodologies, differences in the ancestries of the imputation reference panels, or even different degrees of population stratification within the study populations may have produced the observed differences between our study and that of Hughes et al. The large sample size of our study, our approach to addressing population stratification, and the demonstrated accuracy of our HLA imputation strongly support the conclusion that coding variants account for a major part of the association of *HLA-B*51* with BD.

Associations between other *HLA-B* types and BD have been reported (8–10), as have associations between BD and both *HLA-A* and *HLA-C* alleles (2, 3, 14–17). In our study, we observed that *HLA-B*15, -B*27,* and *-B*57* were independent susceptibility alleles for BD, consistent with earlier reports. We have found suggestive evidence that *HLA-A*26* is an independent risk factor for BD, but, contrary to earlier reports in other populations, neither *HLA-A*02* nor *-A*33* influenced BD risk in our study. Similarly, we found no *HLA-C* region SNP or allele with an *HLA-B*51*–independent effect on BD risk. Interestingly, we have found that *HLA-B*49* and *-A*03* are independent susceptibility alleles that are protective against the development of BD. Although the mechanism through which these alleles protect against BD is unclear, a closer examination of these alleles may provide insights into BD pathogenesis. *HLA-A*03* and *-B*51* are independent alleles that, when coinherited, produce neutral risk of BD (Table S6). HLA-A*03 also is unique among BD-associated molecules in its ability to engage the inhibitory receptor KIR3DL2, and it is possible that enhanced cytotoxic inhibition is protective against BD. The coinheritance of *HLA-B*49* with *HLA-B*51* did not reduce the risk of BD conferred by *HLA-B*51*. It also is worth noting that the HLA-B*49 and -B*51 proteins both contain the Bw4 motif, although differing at four of the five BD-associated amino acid positions.

Through the examination of amino acid sequences from MHC alleles, intramolecular mapping of disease-associated MHC amino acid positions has clarified the MHC associations in both rheumatoid arthritis (19) and host control of HIV infection (31). In BD, we have mapped the association of MHC-I molecules to positions that critically influence the peptide-binding specificity of the MHC molecule, particularly the anchor residues at positions 67 and 116, providing the most definitive data to date implicating peptide binding by MHC-I in BD. The importance of peptide binding is supported further by our recent observation that variants of ERAP1, an important enzyme in the processing of peptides for presentation by MHC-I, confers risk of BD through epistasis with *HLA-B*51* (6). Collectively, these data support a role for one or more pathogenic peptides in BD, vindicating the investigation of infectious agents, exogenous peptides, endogenous autoantigens, and molecular mimicry in BD pathogenesis (33).

In addition to implicating peptide–MHC-I binding in BD, our results suggest that the regulation of CTL and NK activation by MHC-I is involved in the pathogenesis of BD. MHC-I molecules inherently regulate CTL activation by presenting peptide antigens to T-cell receptors; however, many BD-associated MHC-I variants in our study also regulate NK and CTL activation through interactions with the inhibitory KIR3DL1 or KIR3DL2 receptors or with the excitatory KIR3DS1 receptor (22, 24, 29). Furthermore, positions 67 and 116 (and to a lesser extent position 97) of the HLA-B antigen-binding groove are critical determinants of the binding specificity of some HLA-B alleles for KIR3DL1 or KIR3DS1 receptors (22). Beyond the MHC-I antigen-binding groove, we identified an association between BD and a variant of the HLA-B signal peptide that independently regulates CTL and NK-cell activation through HLA-E and the C-type lectin-like heterodimeric receptors, CD94/NKG2 (34–36). Coincidentally, we previously have identified an association between BD and *KLCR4*, which encodes the C-type lectin-like receptor, NKG2F (6).

Although the nature of BD is clearly inflammatory, it has been debated whether this inflammation is autoimmune, resulting from inappropriate adaptive immune activation and broken immune tolerance, or autoinflammatory, resulting from inappropriately activated innate immune pathways. Although the strongest risk factor for BD remains *HLA-B*51*, which encodes what usually is considered an adaptive immune molecule, several recent genetic studies have clearly linked innate immune mechanisms to BD risk, supporting an autoinflammatory contribution to its pathogenesis (6, 7). Our data bring additional insight to this discussion by offering a plausible mechanism through which MHC-I could act to confer risk of BD through the regulation of both innate NK cells and adaptive CTLs.

In summary, we have identified an array of MHC-I alleles that are associated with BD. Within MHC-I molecules, we have identified a group of amino acid positions that strongly influence the risk of developing BD, many of which are located at sites known to influence both peptide binding and the interactions between MHC-I molecules and receptors on CTL and NK cells. Further, our identification of a BD-associated residue in the signal peptide of HLA-B independently links BD pathogenesis to cytotoxicity. Taken together, these data further implicate MHC-I peptide binding in the pathophysiology of BD and provide several lines of evidence independently connecting the regulation of cell-mediated cytotoxicity to BD risk.

Materials and Methods

Patients. BD patients fulfilling the International Study Group Criteria for BD (37) and geographically matched, healthy blood bank control subjects were enrolled in the study at the BD clinic of the Istanbul Faculty of Medicine, Division of Rheumatology, as previously described (2).

SNP Genotyping and Haplotype Analysis. SNP genotype data were generated using Infinium Human CNV370 arrays (Illumina) and were processed with SVS7 (Golden Helix), as previously described (2). The subset of markers

residing within the MHC (n = 2,832) were extracted for subsequent analyses. Haplotypes of SNPs from the *HLA-B/MICA* region with MAF >0.2 and *HLA-B* locus types were assembled manually, and haplotype association testing was performed by χ^2 testing, as implemented in Haploview software (18). When necessary, validation genotyping was performed with custom MassArray assays (Sequenom).

Direct Typing of the *HLA-B* Locus. Two-digit *HLA-B* locus typing of genomic DNA was performed as previously described (2). Associations between BD and *HLA-B* types were assessed using numeric regression in SVS7. Stepwise logistic regression with forward selection and conditional analysis of *HLA-B* types also were performed using SVS7. After correcting for 31 observed HLA-B types, significance was defined as $P < 1.6 \times 10^{-3}$.

SNP Genotype Imputation. Using the 2,832 directly ascertained MHC region SNPs as a foundation, SNP imputation was performed using IMPUTE2 (38) software and using the 1,000 Genomes Project phase 1 integrated dataset as the reference dataset, as described (39). Imputation accuracy was internally assessed by independently masking and imputing every genotyped SNP, from which the overall concordance rate was determined to be 99.6%. Imputed SNP data were filtered to remove rare markers and markers that were imputed with poor quality using SVS7. Association testing and conditional analyses were performed using SNPTESTv2. Significance was determined using the genome-wide significance threshold ($P < 5 \times 10^{-8}$), adjusted for three models ($P < 1.7 \times 10^{-8}$).

Imputation of Classical HLA Types and Amino Acid Residues. The directly ascertained MHC SNPs also were used to impute classical HLA alleles and their corresponding amino acid sequences using SNP2HLA and reference data collected by the Type I Diabetes Genetic Consortium, as described (19). Briefly, this dataset included a panel of 2,537 MHC SNPs that were selected to tag the entire MHC region, together with classical types for eight MHC-I and MHC-II loci at four-digit resolution in 2,767 unrelated individuals of European ancestry. All SNPs, amino acids, and HLA types in the reference panel were encoded as biallelic markers. Multivariate logistic regression analysis of classical HLA types was undertaken 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-I alleles were generated using the EM algorithm, and haplotype association testing was performed with SVS7.

1. Sakane T, Takeno M, Suzuki N, Inaba G (1999) Behçet's disease. *N Engl J Med* 341(17): 1284–1291.
2. 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.
3. Mizuki N, et al. (2010) Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. *Nat Genet* 42(8):703–706.
4. Ono S, Aoki K, Sugiura S, Nakayama E, Itakura K (1973) Letter: HL-A5 and Behçet's disease. *Lancet* 2(7842):1383–1384.
5. Ohno S, et al. (1982) Close association of HLA-Bw51 with Behçet's disease. *Arch Ophthalmol* 100(9):1455–1458.
6. Kirino Y, 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(2): 202–207.
7. Kirino Y, et al. (2013) Targeted resequencing implicates the familial Mediterranean fever gene MEFV and the toll-like receptor 4 gene TLR4 in Behçet disease. *Proc Natl Acad Sci USA* 110(20):8134–8139.
8. Choukri F, Chakib A, Himmich H, Hüe S, Caillat-Zucman S (2001) HLA-B*51 and B*15 alleles confer predisposition to Behçet's disease in Moroccan patients. *Hum Immunol* 62(2):180–185.
9. Gül A, et al. (2002) A weak association of HLA-B*2702 with Behçet's disease. *Genes Immun* 3(6):368–372.
10. Ahmad T, et al. (2003) Mapping the HLA association in Behçet's disease: A role for tumor necrosis factor polymorphisms? *Arthritis Rheum* 48(3):807–813.
11. Mizuki N, et al. (1997) Triplet repeat polymorphism in the transmembrane region of the MICA gene: A strong association of six GCT repetitions with Behçet disease. *Proc Natl Acad Sci USA* 94(4):1298–1303.
12. Ota M, et al. (1999) The critical region for Behçet disease in the human major histocompatibility complex is reduced to a 46-kb segment centromeric of HLA-B, by association analysis using refined microsatellite mapping. *Am J Hum Genet* 64(5): 1406–1410.
13. Mizuki N, et al. (2000) Localization of the pathogenic gene of Behçet's disease by microsatellite analysis of three different populations. *Invest Ophthalmol Vis Sci* 41(12):3702–3708.
14. Hughes T, et al. (2013) Identification of multiple independent susceptibility loci in the HLA region in Behçet's disease. *Nat Genet* 45(3):319–324.
15. Mizuki N, et al. (1997) A strong association between HLA-B*5101 and Behçet's disease in Greek patients. *Tissue Antigens* 50(1):57–60.
16. Meguro A, et al. (2010) Genetics of Behçet disease inside and outside the MHC. *Ann Rheum Dis* 69(4):747–754.
17. Kang EH, et al. (2011) Associations between the HLA-A polymorphism and the clinical manifestations of Behçet's disease. *Arthritis Res Ther* 13(2):R49.
18. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.
19. 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.
20. Sidney J, Peters B, Frahm N, Brander C, Sette A (2008) HLA class I supertypes: A revised and updated classification. *BMC Immunol* 9:1.
21. Huyton T, Ladas N, Schumacher H, Blasczyk R, Bade-Doeding C (2012) Pocketcheck: Updating the HLA class I peptide specificity roadmap. *Tissue Antigens* 80(3):239–248.
22. Sanjanwala B, Draghi M, Norman PJ, Guethlein LA, Parham P (2008) Polymorphic sites away from the Bw4 epitope that affect interaction of Bw4+ HLA-B with KIR3DL1. *J Immunol* 181(9):6293–6300.
23. Borrego F, Ulbrecht M, Weiss EH, Coligan JE, Brooks AG (1998) Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med* 187(5):813–818.
24. Gumperz JE, et al. (1997) Conserved and variable residues within the Bw4 motif of HLA-B make separable contributions to recognition by the NKB1 killer cell-inhibitory receptor. *J Immunol* 158(11):5237–5241.
25. Fadda L, et al. (2011) Common HIV-1 peptide variants mediate differential binding of KIR3DL1 to HLA-Bw4 molecules. *J Virol* 85(12):5970–5974.
26. Malnati MS, et al. (1995) Peptide specificity in the recognition of MHC class I by natural killer cell clones. *Science* 267(5200):1016–1018.
27. Mandelboim O, Wilson SB, Valés-Gómez M, Reyburn HT, Strominger JL (1997) Self and viral peptides can initiate lysis by autologous natural killer cells. *Proc Natl Acad Sci USA* 94(9):4604–4609.
28. Boyington JC, Motyka SA, Schuck P, Brooks AG, Sun PD (2000) Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* 405(6786):537–543.
29. Hansasuta P, et al. (2004) Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. *Eur J Immunol* 34(6):1673–1679.
30. Vivian JP, et al. (2011) Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. *Nature* 479(7373):401–405.
31. Pereyra F, et al.; International HIV Controllers Study (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330(6010): 1551–1557.
32. Dilthey AT, Moutsianas L, Leslie S, McVean G (2011) HLA*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* 27(7):968–972.
33. Direskeneli H (2001) Behçet's disease: Infectious aetiology, new autoantigens, and HLA-B51. *Ann Rheum Dis* 60(11):996–1002.
34. Braud V, Jones EY, McMichael A (1997) The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9. *Eur J Immunol* 27(5):1164–1169.
35. Rodgers JR, Cook RG (2005) MHC class Ib molecules bridge innate and acquired immunity. *Nat Rev Immunol* 5(6):459–471.
36. Merino AM, et al. (2013) Dimorphic HLA-B signal peptides differentially influence HLA-E- and natural killer cell-mediated cytolysis of HIV-1-infected target cells. *Clin Exp Immunol* 174(3):414–423.
37. International Study Group for Behçet's Disease (1990) Criteria for diagnosis of Behçet's disease. *Lancet* 335(8697):1078–1080.
38. 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.
39. Howie B, Marchini J, Stephens M (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1(6):457–70.
40. 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.
41. 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.