Sawcer S, Todd JA, Trowsdale J, Beck S (2004) Complete MHC haplotype sequencing for common disease gene mapping. Genome Res 14:1176–1187

Strichman-Almashanu LZ, Bustin M, Landsman D (2003) Retroposed copies of the HMG genes: a window to genome dynamics. Genome Res 13:800–812

Takasu M, Hayashi R, Maruya E, Ota M, Imura K, Kougo K, Kobayashi C, Saji H, Ishikawa Y, Asai T, Tokunaga K (2007) Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon. Tissue Antigens 70:144–150

Tian W, Wang F, Cai JH, Li LX (2008) Polymorphic insertions in 5 Alu loci within the major histocompatibility complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. Tissue Antigens 72:559–567

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA (2005) SVA elements: a hominid-specific retroposon family. J Mol Biol 354:994–1007

Watanabe Y, Tokunaga K, Geraghty DE, Tadokoro K, Juji T (1997) Large-scale comparative mapping of the MHC class I region of predominant haplotypes in Japanese. Immunogenetics 46:135–141

Witt C, Sayer D, Trimboli F, Saw M, Herrmann R, Cannell P, Baker D, Christiansen F (2000) Unrelated donors selected prospectively by block-matching have superior bone marrow transplant outcome. Hum Immunol 61:85–91

Yao Y, Shi L, Shi L, Lin K, Yu L, Sun H, Huang X, Tao Y, Yi W, Liu S, Chu J (2009) The association between *HLA-A*, -B alleles and major histocompatibility complex class I polymorphic Alu insertions in four populations in China. Tissue Antigens 73:575–581

Genetics of Behçet disease inside and outside the MHC

Akira Meguro,[1] Hidetoshi Inoko,[2] Masao Ota,[3] Yoshihiko Katsuyama,[4] Akira Oka,[2] Eiichi Okada,[5] Ryoji Yamakawa,[6] Takenosuke Yuasa,[7] Toshihiko Fujioka,[8] Shigeaki Ohno,[9] Seiamak Bahram,[10] Nobuhisa Mizuki[1]

► Additional data (supplementary figures and tables) are published online only at http://ard.bmj.com/content/vol69/issue4

[1]Department of Ophthalmology and Visual Science, Yokohama City University Graduate School of Medicine, Kanazawa-ku, Yokohama, Kanagawa, Japan
[2]Department of Molecular Life Science, Division of Molecular Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Kanagawa, Japan
[3]Department of Legal Medicine, Shinshu University School of Medicine, Matsumoto, Nagano, Japan
[4]Department of Pharmacy, Shinshu University School of Medicine, Matsumoto, Nagano, Japan
[5]Okada Eye Clinic, Kounan-ku, Yokohama, Kanagawa, Japan
[6]Department of Ophthalmology, Kurume University School of Medicine, Kurume, Fukuoka, Japan
[7]Yuasa Eye Clinic, Nishi-ku, Osaka, Osaka, Japan
[8]Fujioka Eye Hospital, Hakodate, Hokkaido, Japan
[9]Department of Ocular Inflammation and Immunology, Hokkaido University Graduate School of Medicine, Sapporo, Hokkaido, Japan
[10]Human Molecular Immunogenetics, Centre de Recherche d'Immunologie et d'Hématologie, School of Medicine, Strasbourg Cedex, France

Correspondence to
Professor Seiamak Bahram, Strasbourg School of Medicine, 4 rue Kirschleger, Strasbourg 67085, France; siamak@hemato-ulp.u-strasbg.fr

Accepted 23 July 2009
Published Online First 13 August 2009

ABSTRACT

Background Behçet disease (BD) is a rare, chronic, systemic, inflammatory disorder characterised by recurrent ocular, genital and skin lesions. Although its aetiology is still uncertain, an intricate interplay between the environment (eg, viruses) and the host seems to initiate and/or perpetuate the disease, although the mechanism remains speculative. Since the identification of *HLA-B*5101* (and more recently of *MICA*) as a susceptibility locus for BD, the identification of additional genetic locus/loci, whether inside, or perhaps more importantly outside the MHC has clearly stalled.

Objective To carry out a genome-wide association study (GWAS) of BD.

Methods 300 Japanese patients with BD and an equal number of controls were recruited. The samples were screened using a dense panel of 23 465 microsatellites (MS) covering the entire genome.

Results The six best (of a total of 147) positively associated MS with BD were identified. Of these six, two were located within the human leucocyte antigen (HLA) class I region itself. Although one of these was clearly reminiscent of the association with *HLA-B,* the second, not in linkage disequilibrium with the former, was in the telomeric side of the class I region and remained to be formally identified. HLA genotyping and haplotype analysis conclusively led to the deciphering of a dual, independent, contribution of two HLA alleles to the pathogenesis of BD: *HLA-B*5101* and *HLA-A*26*.

Conclusions This GWAS highlights the premier genetic susceptibility locus for BD as the major histocompatibility complex itself, wherein reside two independent loci: *HLA-B* and *HLA-A*.

INTRODUCTION

Behçet disease (BD) is a rare, chronic, systemic, inflammatory disorder characterised by four major symptoms consisting of recurrent ocular symptoms, oral and genital ulcers and skin lesions.[1] BD exists world wide but is clearly more prevalent in countries along the ancient silk route, spanning from Japan to the Middle East and the Mediterranean.[2]

The aetiology of BD remains uncertain, but as in many other inflammatory and/or immune-centred diseases, environmental factors are thought to trigger the pathology in people with a 'favourable' genetic background. The sole agreed genetic component of BD, persistent across the globe, is the human leucocyte antigen (HLA) class I antigen, *HLA-B51*.[2] It remains to be firmly and formally established—as is the case of the quasi-totality of HLA-linked diseases—whether this HLA allele itself or a closely linked gene(s) is the contributing genetic factor. We and others have shown that *HLA-B*5101* itself, alone[3–5] or in conjunction with the closely linked *MICA*009* allele, appears to be the main major histocompatibility complex (MHC)-based contributing factor in the development of BD.[5–7]

The MHC is important for any insight into the genetics of diseases having an expansive immune component. Until recently, it has been difficult to venture beyond the MHC, as the tools to do so were not available. There are three main approaches to tracking the genetic basis of diseases carrying an inherited component: linkage studies, which are suitable for Mendelian disorders; genome-wide association studies (GWAS) for complex diseases and, finally, the candidate gene approach, which can be used on its own or within the settings of the first two approaches. As applied to BD, the candidate gene approach has allowed the identification—besides *HLA-B51*—of a number of disparate loci,[8–10] but most of the studies had a small sample size and a lack of replication studies. Although, non-Mendelian diseases are not well suited for linkage studies, Karasneh *et al* completed such an analysis in BD using a set of 428 microsatellites (MS) in Turkish multicase families.[11] The study reported the identification of several non-HLA susceptibility loci for the disease. However, the non-parametric linkage analysis performed in the study had several inherent weaknesses, including poor mapping resolution and low detection power. Finally, GWAS is appropriate for a disease such as BD,[12] and hence we embarked on the first GWAS in BD.

MATERIALS AND METHODS

Patients and healthy controls

Detailed information about patients, controls and methods is presented in an online supplementary text file. In brief, 300 unrelated patients with BD and 300 unrelated healthy controls, all of Japanese descent, were enrolled in this study. The diagnosis of BD was established according to standard criteria[13] proposed by the Japan Behçet's Disease Research Committee. The patients and controls were equally distributed in groups of 100 vs 100 in each and all of the three pooled DNA typing stages (online supplementary table 1). After pooled DNA typing, individual DNA typing was performed on the same sample set. The study methodology complied with the guidelines of the Declaration of

Table 1 Six positive microsatellite markers in the pooled DNA screening and in the individual DNA screening

		p Value					
		Pooled DNA (2×2)			Individual DNA		
Marker	Chromosome	First	Second	Third	2×2	2×m	Nearest gene
D3S0186i	3p12	<0.0001	<0.0001	<0.0001	0.029	0.15	*ROBO1*
D6S0014i	6p21.3	0.002	0.0003	0.003	<0.0001	<0.0001	*FLJ45422*
D6S0032i	6p21.3	<0.0001	0.0002	0.0005	<0.0001	<0.0001	*HLA-B*
536G12A	6q25.1	<0.0001	<0.0001	<0.0001	0.031	0.111	*PPIL4*
D12S0645i	12p12.1	<0.0001	<0.0001	<0.0001	0.024	0.364	*SOX5*
D22S0104i	22q11.22	<0.0001	<0.0001	0.006	0.019	0.125	*IGL@*

Table 2 Haplotype frequencies of *HLA-A, -F* and *-G* in patients with Behçet disease (BD) and controls

	Frequency				
Haplotype	BD	Controls	OR (95% CI)	p	Pc
All subjects (BD:Controls=300:300)					
*A*0201-F*010101-G*010101*	0.096	0.086	1.12 (0.75 to 1.65)		
*A*0206-F*010101-G*010101*	0.071	0.078	0.91 (0.59 to 1.39)		
*A*1101-F*010103-G*010401*	0.005	0.020	0.23 (0.06 to 0.86)	0.034	NS
*A*2402-F*010102-G*010401*	0.067	0.133	0.47 (0.31 to 0.69)	0.00016	0.0086
*A*2402-F*010103-G*010401*	0.254	0.243	1.06 (0.82 to 1.38)		
*A*2601-F*010101-G*010102*	0.130	0.070	1.99 (1.34 to 2.96)	0.00051	0.028
*A*2602-F*010101-G*010102*	0.030	0.015	2.04 (0.91 to 4.58)		
*A*2603-F*010101-G*010102*	0.033	0.018	1.82 (0.86 to 3.85)		
*A*2605-F*010101-G*010102*	0.003	0.000	—		
**3101-F*010103-G*010102*	0.107	0.064	1.75 (1.16 to 2.66)	0.0071	NS
*A*3101-F*010103-G*0103*	0.010	0.000	—	0.015	NS
*A*3303-F*010102-G*010401*	0.025	0.072	0.33 (0.18 to 0.61)	0.00022	0.012
*A*26-F*010101-G*010102*	0.197	0.103	2.13 (1.53 to 2.97)	5.34×10^{-6}	0.00028
*HLA-B*5101* non-carriers (BD:Controls=136:255)					
*A*0201-F*010101-G*010101*	0.104	0.085	1.25 (0.76 to 2.07)		
*A*0201-F*010103-G*010101*	0.026	0.006	4.48 (1.15 to 17.47)	0.038	NS
*A*0206-F*010101-G*010101*	0.063	0.075	0.83 (0.46 to 1.50)		
*A*2402-F*010102-G*010401*	0.070	0.146	0.44 (0.26 to 0.75)	0.0017	NS
*A*2402-F*010103-G*010401*	0.256	0.244	1.07 (0.76 to 1.50)		
*A*2601-F*010101-G*010102*	0.174	0.073	2.68 (1.69 to 4.25)	2.89×10^{-5}	0.0013
*A*2602-F*010101-G*010102*	0.048	0.014	3.62 (1.43 to 9.19)	0.0071	NS
*A*2603-F*010101-G*010102*	0.033	0.014	2.47 (0.91 to 6.70)		
*A*3303-F*010102-G*010401*	0.030	0.083	0.34 (0.16 to 0.73)	0.0033	NS
*A*26-F*010101-G*010102*	0.256	0.100	3.08 (2.07 to 4.58)	2.30×10^{-8}	9.88×10^{-7}

Only haplotypes that reached frequencies >5% in patients with BD or controls or that may have an effect on the development of BD are listed.
OR, odds ratio; Pc, corrected p value.

Helsinki. The study details were explained to all patients and controls before obtaining their consent to genetic screening.

GWAS

Pooled DNA construction

The protocol of Collins *et al*,[14] with slight modifications,[15] was used to perform MS typing on pooled DNA.[14]

Genome-wide MS genotyping

All MS markers and the genotyping methods for MS analysis used in this study have been previously described.[16]

HLA genotyping and haplotype analysis

Twenty MS markers covering the entire HLA class I region were used in this study. Genotyping of *HLA-A* and *HLA-B* genes was performed using the PCR-sequence-specific oligonucleotide probes-Luminex method.[17] The genotyping of *HLA-E, HLA-F*

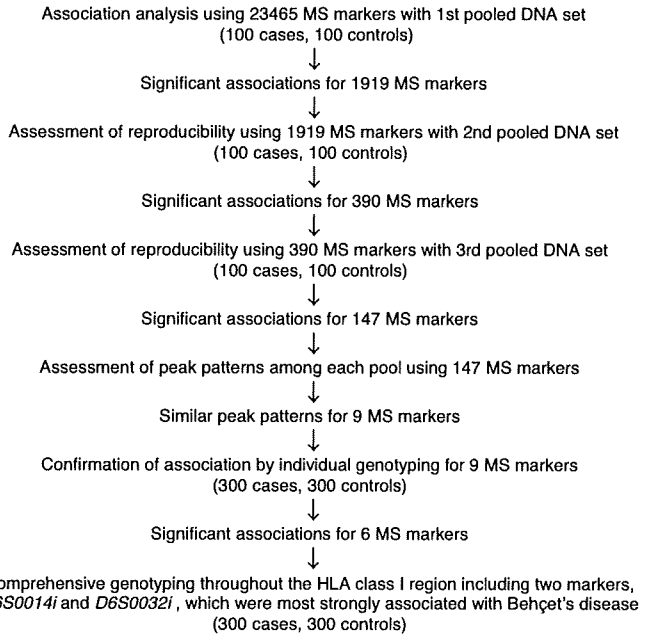

Figure 1 Flow chart of this genome-wide association study of Behçet disease with 23 465 microsatellite markers in 300 cases and 300 controls.

and *HLA-G* genes was performed by direct DNA sequencing (supplementary table 2).[18–20]

RESULTS

Genome-wide association study

The genetic basis of BD remains to a great extent untapped, once the HLA locus is discarded. In this study, we present the first GWAS of BD achieved through three rounds of pooled DNA screening using a comprehensive set of 23 465 MS markers. In the first screen, using 23 465 markers, we found a significant association for 1919 markers, as assessed by the Fisher exact test. In the second screen, 390 of these 1919 markers continued to show a significant association, but after the third screen, only 147 markers remained positive (figure 1, supplementary figure 1, supplementary table 3). A total of nine of these 147 markers had similar peak patterns among the first, second and third case and control pools. To confirm, beyond doubt, the associations observed, we individually genotyped each of these nine markers in the same set of 600 screened subjects. Six markers remained positive after this individual screening step (table 1, figure 1, supplementary figure 1). Among them two—*D6S0014i* and *D6S0032i*—were located in the HLA class I region (6p21.3)—~1.1 Mb and ~36 kb telomeric of the *HLA-B* gene, respectively—and were indeed most strongly associated with BD.

HLA genotyping and haplotype analysis

The location of the two most strongly BD-associated markers in the HLA class I region refocused our interest on the MHC. The HLA class I region spans a region of 1.8 Mb from the centromeric *MICB* to the telomeric *HLA-F* and contains in addition to these two 'gatekeepers', the following MHC-I and related molecules (in the same centro-telomeric direction): *MICA, HLA-B, HLA-C, HLA-E, HLA-A* and *HLA-G*, in addition to a number of non-HLA related genes. Given the physical location of the two MS markers, the association with *D6S0032i* might indeed be reminiscent of that with *HLA-B* and that with *D6S0014i*, reflecting a yet to be identified HLA or non-HLA gene(s). Previous studies have reported the association of BD with certain *HLA-A* alleles in some ethnic groups.[21–23] Additionally, Park *et al* showed that *HLA-E* and *HLA-G* polymorphisms contribute to the development of BD in Korea.[24] However, to date, it remains to be shown whether the association between each gene (*HLA-A, HLA-E* and *HLA-G*) and BD is a primary one or subsequent to linkage disequilibrium (LD) with other genes, including *HLA-B*. In order to identify the genuine MHC-linked pathogenic BD locus/loci we had to circumvent the LD across the region dictated by the only polymorphic loci tested to date—that is, HLA genes. To do so we performed a comprehensive MS mapping and HLA genotyping throughout the HLA class I region. Association, LD and haplotype analyses were conducted using these data. To our knowledge, this is the first report of a genetic association study for BD that focuses on the entire ~2.0 Mb HLA class I region.

A total of 20 MS and five *HLA* genes in and around the HLA class I region were used to genotype the 300 patients and 300 controls (figure 2A,B). For 22/25 genes and MS, statistically significant differences in allele frequencies were found between patients and controls. Among these, *HLA-B* ($Pc=1.63\times10^{-24}$), *D6S2938* ($Pc=5.75\times10^{-12}$) and *D6S2930* ($Pc=1.15\times10^{-13}$) were most strongly associated with BD. In addition, genes and MS in and around the region from *HLA-F* to *HLA-G* were markedly associated with BD (*HLA-F*, $Pc=1.83\times10^{-8}$; *HLA-G*, $Pc=1.44\times10^{-6}$; *174B01*, $Pc=9.91\times10^{-8}$; *180G06*, $Pc=6.55\times10^{-7}$; *D6S2707*, $Pc=8.37\times10^{-7}$). The frequencies of the identified *HLA-A, -B, -E, -F* and *-G* alleles are listed in supplemental table 4.

*HLA-B*5101* was the strongest susceptibility allele for the development of BD (odds ratio (OR)=5.50). In other HLA genes, *HLA-A*2601* (OR=1.92, Pc=0.016), *HLA-F*010101* (OR=1.43, Pc=0.011) and *HLA-G*010102* (OR=2.08, $Pc=1.44\times10^{-6}$) were positively associated with susceptibility to BD. Conversely, *HLA-A*3303* (OR=0.34, Pc=0.0035), *HLA-F*010102* (OR=0.41, $Pc=1.83\times10^{-8}$) and *HLA-G*010401* (OR=0.41, $Pc=3.67\times10^{-4}$) were negatively associated with susceptibility to BD. No association with *HLA-E* was found.

A sizeable fraction of patients with BD do not carry the *HLA-B*5101* allele. What is their genetic makeup? To examine this, we stratified the patient and control groups according to the presence or absence of *HLA-B*5101*. A total of 136 patients with BD and 255 controls did not carry *HLA-B*5101* (figure 2C, supplemental table 4). No other *HLA-B* alleles were associated with BD. Three MS located within a region 260 kb telomeric of *HLA-B* (*D6S2938, D6S2930* and *D6S0032i*) were only slightly associated with BD. In contrast, a strong association between BD and a 770 kb region including *HLA-A, -F* and *-G* (from *D6S2770* to *D6S0014i*) was observed, hence independently of *HLA-B*5101*. *D6S2838*, located between *HLA-A* and *HLA-G*, was most strongly associated with BD (OR=2.75, $Pc=5.98\times10^{-7}$). *HLA-A*2601* (OR=2.61, $Pc=5.11\times10^{-4}$), *HLA-F*010101* (OR=1.87, $Pc=1.55\times10^{-4}$) and *HLA-G*010102* (OR=2.24, $Pc=1.16\times10^{-4}$) were positively associated with susceptibility to BD, whereas *HLA-F*010102* (OR=0.40, $Pc=2.09\times10^{-5}$) and *HLA-G*010401* (OR=0.56, $Pc=2.41\times10^{-3}$) were negatively associated with susceptibility to BD.

Figure 3 shows the overall LD patterns for the 20 MS and five HLA genes in 300 patients with BD and 300 controls. Strong LD was seen throughout the ~380 kb region including *HLA-A, HLA-F* and *HLA-G* (from 186A07 to 022G02), and strong associations were observed for *HLA-A-HLA-G* (D' of cases and controls ($D'_{cases\&controls}$)=0.940 and 0.902, respectively), *HLA-A-HLA-F* ($D'_{cases\&controls}$=0.789 and 0.718) and *HLA-F-HLA-G* ($D'_{cases\&controls}$=0.662 and 0.746). *D6S0032i*, one of two positive MS markers located in the HLA class I region identified by our genome-wide case–control study, was most strongly associated

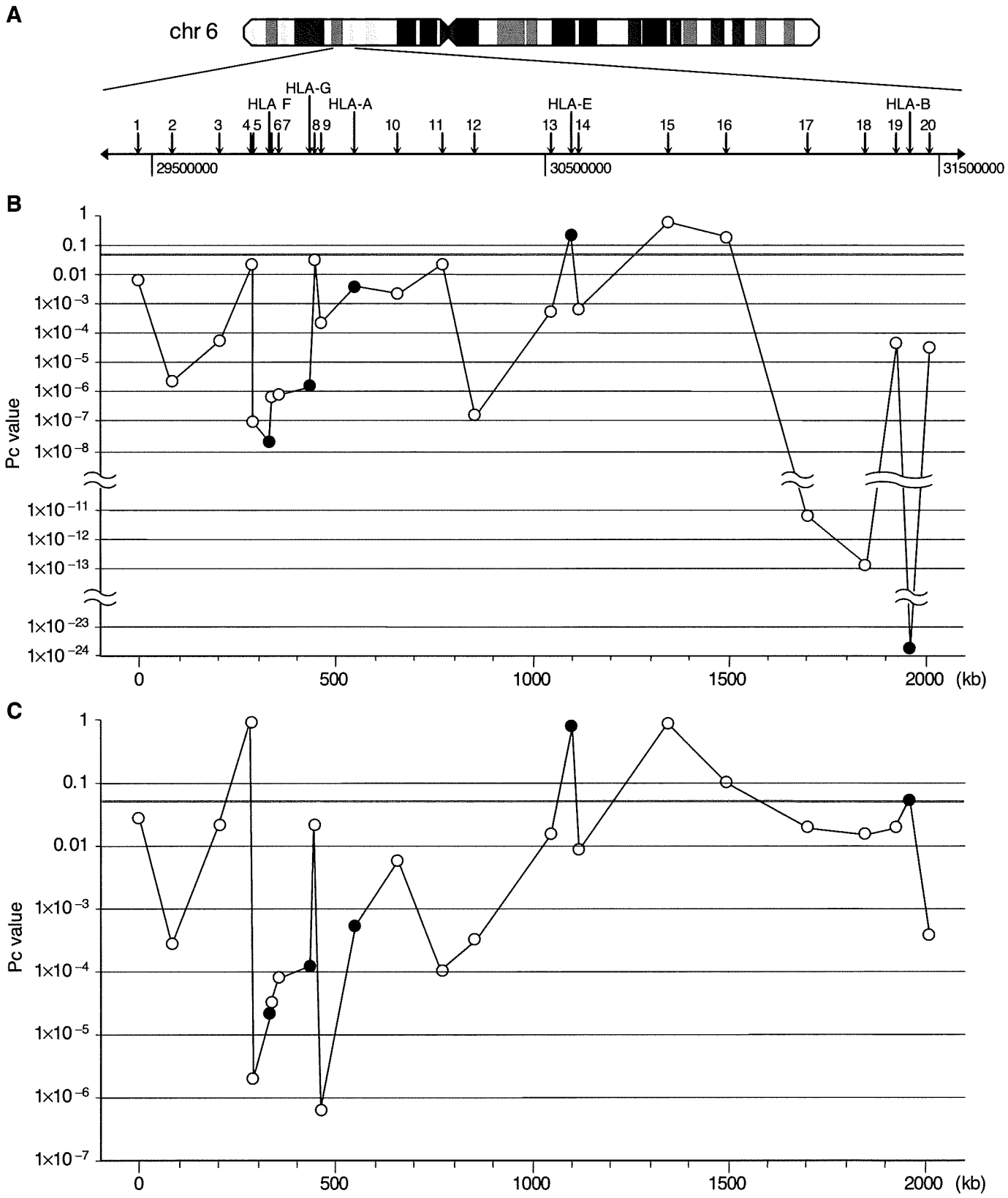

Figure 2 Allele differences in HLA genes and microsatellites (MS) between Japanese patients with Behçet disease (BD) and healthy controls. (A) Diagram of chromosome 6 showing the locations of HLA genes and MS examined in this study. MS are indicated by the following numbers: 1, D6S0016i; 2, D6S2770; 3, D6S2910; 4, D6S0506i; 5, D6S0129i; 6, D6S0350i; 7, D6S2707; 8, D6S2837; 9, D6S2838; 10, D6S265; 11, D6S1103i; 12, D6S0014i; 13, D6S2840; 14, D6S2799; 15, D6S2956; 16, D6S2825; 17, D6S2938; 18, D6S2930; 19, D6S0032i; 20, MICA-TM. (B) Allele differences between the Japanese patients with BD and control groups in the HLA genes and MS used for association analysis. (C) Allele differences between the Japanese HLA-B*5101 non-carrier patients with BD and control groups in the HLA genes and MS used for association analysis. The vertical and horizontal axes are the corrected P (Pc) value and the location of genes and MS in the studied 2.0 Mb region, respectively. Black and white circles denote genes and MS, respectively. The red line shows the significance threshold (ie, Pc=0.05).

D'	1	2	3	4	5	F	6	7	G	8	9	A	10	11	12	13	E	14	15	16	17	18	19	B	20
1	-	0.704	0.533	0.458	0.420	0.392	0.377	0.520	0.330	0.204	0.148	0.429	0.288	0.274	0.298	0.378	0.200	0.158	0.163	0.280	0.309	0.379	0.307	0.491	0.170
2	0.719	-	0.648	0.950	0.911	0.890	0.915	0.920	0.900	0.831	0.982	0.682	0.769	0.530	0.970	0.664	0.876	0.476	0.602	0.709	0.767	0.716	0.538	0.813	0.816
3	0.527	0.605	-	0.545	0.642	0.568	0.439	0.744	0.357	0.237	0.584	0.557	0.620	0.460	0.421	0.434	0.531	0.704	0.416	0.413	0.350	0.387	0.359	0.504	0.249
4	0.549	0.958	0.507	-	0.897	0.789	0.770	0.735	0.514	0.439	0.975	0.672	0.484	0.351	0.284	0.469	0.412	0.376	0.226	0.296	0.320	0.364	0.251	0.437	0.205
5	0.537	0.976	0.581	0.915	-	0.877	0.809	0.750	0.719	0.524	0.971	0.774	0.640	0.422	0.507	0.537	0.388	0.401	0.312	0.356	0.394	0.364	0.308	0.459	0.229
F	0.462	0.966	0.600	0.743	0.876	-	0.941	0.830	0.662	0.607	0.972	0.789	0.696	0.393	0.215	0.484	0.345	0.147	0.236	0.246	0.277	0.355	0.285	0.411	0.233
6	0.458	0.970	0.526	0.736	0.850	0.905	-	0.806	0.648	0.590	0.828	0.776	0.680	0.369	0.328	0.501	0.365	0.216	0.291	0.308	0.301	0.382	0.273	0.418	0.262
7	0.554	0.969	0.771	0.774	0.828	0.882	0.865	-	0.817	0.648	0.783	0.774	0.696	0.391	0.538	0.536	0.440	0.373	0.353	0.362	0.411	0.415	0.324	0.501	0.375
G	0.315	0.958	0.355	0.439	0.751	0.746	0.751	0.845	-	0.751	0.987	0.940	0.730	0.349	0.557	0.464	0.167	0.317	0.196	0.321	0.246	0.331	0.173	0.396	0.207
8	0.308	0.955	0.416	0.377	0.608	0.693	0.674	0.730	0.822	-	0.575	0.784	0.701	0.373	0.339	0.419	0.175	0.146	0.207	0.244	0.288	0.334	0.205	0.384	0.220
9	0.256	1.000	0.707	0.920	0.640	0.957	0.615	0.708	0.856	0.680	-	0.990	0.725	0.584	0.654	0.575	0.501	0.547	0.273	0.242	0.485	0.308	0.172	0.370	0.301
A	0.456	0.780	0.600	0.594	0.729	0.718	0.769	0.769	0.902	0.844	0.988	-	0.867	0.458	0.604	0.565	0.342	0.350	0.322	0.372	0.388	0.388	0.312	0.466	0.312
10	0.389	0.763	0.618	0.421	0.637	0.665	0.680	0.680	0.754	0.738	0.741	0.860	-	0.457	0.436	0.543	0.181	0.229	0.210	0.371	0.288	0.356	0.337	0.412	0.285
11	0.299	0.528	0.480	0.366	0.427	0.443	0.437	0.396	0.392	0.403	0.581	0.469	0.464	-	0.333	0.503	0.375	0.382	0.254	0.271	0.335	0.329	0.280	0.371	0.191
12	0.421	0.939	0.610	0.227	0.538	0.433	0.424	0.511	0.491	0.427	0.606	0.532	0.481	0.364	-	0.518	0.321	0.295	0.176	0.219	0.267	0.289	0.203	0.404	0.086
13	0.440	0.783	0.594	0.478	0.534	0.418	0.475	0.554	0.391	0.377	0.579	0.526	0.527	0.442	0.449	-	0.545	0.548	0.390	0.459	0.398	0.457	0.360	0.552	0.385
E	0.302	0.896	0.340	0.276	0.444	0.332	0.353	0.470	0.337	0.321	0.216	0.371	0.249	0.348	0.307	0.587	-	0.584	0.279	0.270	0.346	0.351	0.265	0.482	0.196
14	0.284	0.474	0.648	0.272	0.342	0.197	0.216	0.322	0.251	0.198	0.423	0.338	0.243	0.257	0.269	0.560	0.608	-	0.328	0.423	0.179	0.279	0.176	0.439	0.129
15	0.296	0.839	0.299	0.301	0.346	0.332	0.382	0.437	0.232	0.256	0.124	0.319	0.248	0.352	0.134	0.457	0.323	0.313	-	0.263	0.279	0.296	0.189	0.438	0.178
16	0.363	0.753	0.523	0.339	0.375	0.266	0.344	0.414	0.307	0.323	0.324	0.455	0.389	0.348	0.193	0.499	0.386	0.494	0.354	-	0.481	0.388	0.412	0.559	0.459
17	0.399	0.824	0.407	0.398	0.495	0.478	0.510	0.552	0.364	0.415	0.559	0.451	0.341	0.341	0.309	0.508	0.441	0.344	0.407	0.488	-	0.662	0.555	0.765	0.656
18	0.432	0.765	0.371	0.357	0.386	0.327	0.387	0.446	0.366	0.354	0.220	0.423	0.363	0.352	0.277	0.463	0.356	0.337	0.443	0.424	0.624	-	0.534	0.747	0.591
19	0.324	0.583	0.407	0.226	0.336	0.292	0.351	0.368	0.257	0.272	0.336	0.384	0.292	0.382	0.177	0.379	0.249	0.301	0.296	0.472	0.487	0.570	-	0.879	0.543
B	0.530	0.844	0.542	0.480	0.552	0.474	0.514	0.584	0.472	0.479	0.466	0.549	0.512	0.497	0.485	0.612	0.510	0.499	0.506	0.607	0.820	0.779	0.833	-	0.944
20	0.289	0.794	0.288	0.252	0.351	0.342	0.369	0.391	0.344	0.362	0.340	0.331	0.283	0.172	0.223	0.411	0.366	0.232	0.364	0.490	0.681	0.578	0.470	0.932	-

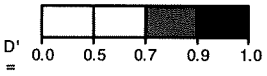

Figure 3 Linkage disequilibrium (LD) values between each microsatellite (MS)/gene in patients with Behçet disease and controls. LD measures are presented in *D´*, a standardised measure that ranges from 0 to 1. *D´* in patients and controls are shown at the top right and the bottom left, respectively. MS and genes are indicated by the following symbols: 1, D6S0016i; 2, D6S2770; 3, D6S2910; 4, D6S0506i; 5, D6S0129i; 6, D6S0350i; 7, D6S2707; 8, D6S2837; 9, D6S2838; 10, D6S265; 11, D6S1103i; 12, D6S0014i; 13, D6S2840; 14, D6S2799; 15, D6S2956; 16, D6S2825; 17, D6S2938; 18, D6S2930; 19, D6S0032i; 20, MICA-TM; A, HLA-A; B, HLA-B; E, HLA-E; F, HLA-F; G, HLA-G.

with *HLA-B* ($D'_{cases\&controls}$=0.879 and 0.833). *HLA-B* exhibited a strong LD with *MICA-TM* (transmembrane) polymorphism ($D'_{cases\&controls}$=0.944 and 0.932), *D6S2938* ($D'_{cases\&controls}$=0.765 and 0.820) and *D6S2930* ($D'_{cases\&controls}$=0.747 and 0.779). On the other hand, *D6S0014i*, another positive marker identified by our genome-wide study, was more strongly associated with *HLA-A* ($D'_{cases\&controls}$=0.604 and 0.532) than with other HLA genes and exhibited the strongest LD with *D6S2770* ($D'_{cases\&controls}$=0.970 and 0.939). Interestingly, *D6S2770* exhibited a strong LD with most MS and with all HLA genes throughout the ~2.0 Mb region.

Table 2 shows the frequencies of *HLA-A-HLA-F-HLA-G* haplotypes in 300 patients with BD and 300 controls. The frequency of the *HLA-A*2601-F*010101-G*010102* haplotype consisting of three alleles positively associated with BD susceptibility was significantly increased in patients with BD (OR=1.99, Pc=0.028). Furthermore, the frequency of haplotypes containing *HLA-F*010101*, *HLA-G*010102* and other *HLA-A*26* subtypes was increased in patients with BD, but not significantly. The frequency of the haplotype consisting of *HLA-A*26*, *HLA-F*010101* and *HLA-G*010102* was more significantly increased in patients with BD (OR=2.13, Pc=2.77×10^{-4}). In contrast, the haplotype frequencies of *HLA-A*2402-F*010102-G*0101401* and *HLA-A*3303-F*010102-G*0101401* were significantly decreased in patients with BD (OR=0.47, Pc=8.60×10^{-3}; OR=0.33, Pc=0.012, respectively). Table 2 also shows the frequencies of *HLA-A–HLA-F–HLA-G* haplotypes in *HLA-B*5101* non-carriers, 136 patients with BD and 255 controls. The frequency of the *HLA-A*2601-F*010101-G*010102* haplotype was significantly increased in patients with BD (OR=2.68, Pc=1.30×10^{-3}). Moreover, the *HLA-A*26-F*010101-G*010102* haplotype was most strongly associated with susceptibility to BD (OR=3.08, Pc=9.88×10^{-7}).

DISCUSSION

In this study, we present the first GWAS of BD. Our GWAS employed 23 465 highly polymorphic MS markers to cover the euchromatic area (~90%) of the human genome at average intervals of 115.1 kb. Among the ultimate six positive markers which passed several filters, two—*D6S0014i* and *D6S0032i*—were located within the HLA class I region, respectively ~36 kb and ~1.1 Mb telomeric of *HLA-B*. Two other positive markers, *536G12A* and *D12S0645i*, were located on 6q25.1 and 12p12.1, respectively. These two chromosome regions were consistent with susceptibility regions, 6q25-26 and 12p12-13, found by Karasneh *et al*[11] by whole-genome linkage analysis using multicase families. The closest gene to *536G12A* is *PPIL4* (peptidylprolyl isomerase (cyclophilin)-like 4). *PPIL4* is a member of the cyclophilin family, a highly conserved family of proteins that exhibit peptidylprolyl *cis–trans* isomerase activity, and have a role in protein folding, immunosuppression by ciclosporin A and

infection of HIV-1 virions as well as lymphocyte apoptosis.[25–29] *D12S0645i* is located in intron 3 of the *SOX5* (SRY (sex-determining region Y)-box 5) gene. *SOX5* is a member of the human *SOX* gene family; genes with critical roles in a number of embryonic developmental processes.[30–35] *D3S0186i* and *D22S0104i* are located on 3p12 and 22q11.22, respectively. *D3S0186i* is located in intron 3 of the *ROBO1* (roundabout, axon guidance receptor, homologue 1) gene. ROBO1, an immunoglobulin superfamily member, encodes an axon guidance receptor and a cell adhesion receptor.[36] *D22S0104i* is located in the immunoglobulin λ (*IGL*) locus, containing a total of 88 *IGL* genes of which 39–43 are functional. *IGL* directs the synthesis of λ-type light chains and has an important role in the antibody response.[37] There are 18 *IGL* genes within 100 kb from *D22S0104i*, and the nearest gene is *IGLV1-40* (IGL variable 1–40) located 1.4 kb centromeric of *D22S0104i*. Previous studies have shown that the *IGL* locus or genes polymorphisms are associated with autoimmune disorders.[38 39] In addition, an association between BD and elevations of some immunoglobulins has been noted.[40–43]

The intra-MHC positive markers did merit further exploration. We investigated polymorphisms of 20 MS and five HLA genes in the HLA class I region and performed LD and haplotype analyses. We found that *HLA-B*5101* and MS near the *HLA-B* gene were most strongly associated with BD. In addition, the *HLA-A*, *HLA-F* and *HLA-G* alleles and MS in the vicinities of these genes were also associated with BD. After stratification of subjects by their *HLA-B*5101* status and examining the *HLA-B*5101* non-carriers, *HLA-A*, *-F* and *-G* alleles and MS near these genes remained strongly associated with susceptibility to BD. Hence, two independent, LD-strong segments at both ends of the MHC class I region—that is, respectively centred around *HLA-B* and a 380 kb region including *HLA-A*, *HLA-F* and *HLA-G*, were positively associated with BD. It is assumed that *HLA-B*5101* is one of such locus; the other being embedded within the 380 kb region harbouring to date four expressed genes: *HLA-A*, *HLA-F*, *HLA-G* and *HCG9*.[44]

Although polymorphism of *HCG9*, a gene of unknown functional relevance, was not analysed in this study, we showed that *HLA-A*, *HLA-F* and *HLA-G* alleles were significantly associated with BD. In *HLA-F*, *HLA-F*010101* was positively associated with susceptibility to BD, whereas *HLA-F*010102* was negatively associated with susceptibility to BD. The single nucleotide polymorphism that distinguishes between *HLA-F*010101* and *HLA-F*010102* results in a synonymous substitution, and hence is of no functional consequence. Therefore, if *HLA-F*010101* is the primary pathogenic gene that increases the risk for development of BD, *HLA-F*010102*, which has the same amino acid sequence, should also be positively associated with the risk for BD. However, these alleles had opposite effects. Furthermore, although *HLA-G*010102* conferred an increased risk of BD, the incidences of other *HLA-G*0101* alleles (*HLA-G*010101*, *-G*010103*, *-G*010105*, *-G*010107* and *-G*010108*) were not similarly increased in patients with BD. These results therefore suggest that the association with *HLA-F* and *HLA-G* alleles was due to LD—with a genuine, to be identified, pathogenic gene—and not a primary association.

HLA-F and *HLA-G* are located close to *HLA-A* where *HLA-A*2601* was positively associated with susceptibility to BD. Moreover, there is a strong LD between *HLA-F*/-G* and *HLA-A*. In addition, the haplotype consisting of *HLA-A*2601*, *HLA-F*010101* and *HLA-G*010102* was significantly increased in patients with BD. Haplotypes consisting of other *HLA-A*26* subtypes, *HLA-F*010101* and *HLA-G*010102*, were also increased in patients with BD, although not significantly. Therefore, the significantly increased incidences of *HLA-F*010101* and *HLA-G*010102* in patients with BD probably result from LD with *HLA-A*26* subtypes, and it is therefore suggested that *HLA-A*26* represents an independent susceptibility allele for BD in the HLA class I region. In fact, the allele and phenotype frequencies of *HLA-A*26*, including all *HLA-A*26* subtypes, were strongly associated with BD. Especially, among the *HLA-B*5101* non-carriers of this study, the allele and phenotype frequencies of *HLA-A*26* were the most strongly associated with susceptibility to BD of all analysed alleles (allele: 25.9% vs 10.6%, OR=2.96, $Pc=3.63\times10^{-7}$; phenotype: 48.5% vs 19.6%, OR=3.86, $Pc=3.20\times10^{-8}$) (table 3).

In previous studies, we and other groups showed that BD is significantly associated with *HLA-A*26* in Japanese, Taiwanese and Greek populations.[21–23] In addition, it has been reported that the phenotype frequency of *HLA-A*26* was increased about seven times in Saudi Arabian patients compared with healthy controls, but this difference was not significant.[45] However, the role (primary or secondary) of *HLA-A*26* in the development of BD remained unclear because these studies did not include comprehensive analysis of the HLA class I region, and the genotyping was not performed on large samples. In other words these studies were unable to show if the association with the *HLA-A* locus was a primary one or secondary to association with *HLA-B*.

*HLA-A*26* is one of the most common HLA alleles in Asia with at least 38 distinct subtypes (http://www.ebi.ac.uk/imgt/hla/align.html, accessed 18 January 2010). Moreover this allele has been documented to be involved in a variety of pathophysiological conditions.[46–51] Our findings here confirm the results of previous studies and clarify that the *HLA-A*26* allele itself is a primary susceptibility gene involved in the development of BD. Park *et al*[24] have reported that the *HLA-G*010102* allele was significantly increased in Korean patients with BD. However, judging from the LD analysis in this study, it is likely that the significantly increased incidence of *HLA-G*010102* in patients with BD resulted secondarily from a strong LD with *HLA-A*26*, and *HLA-A*26* itself is significantly increased in Korean patients with BD. On the other hand, the phenotype frequency of *HLA-A*26* was slightly increased in Palestinian, Jordanian and Iranian patients with BD compared with healthy controls.[52 53] Furthermore, in Ireland, Italy and Turkey, the phenotype frequency of *HLA-A*26* was increased in healthy controls compared with patients with BD, but not significantly.[54–56]

It is not clear why this association has not been found in all populations. Several reasons can be suggested: (*a*) *HLA-A*26* is more common in Asia, especially Japan and Taiwan, than in other areas. Therefore we could easily find the association between *HLA-A*26* and BD in Japan and Taiwan. (*b*) Previous studies did not recruit enough samples to provide significant association results and did not perform stratification according to *HLA-B*51* status; hence they could not detect the association between *HLA-A*26* and BD in these populations. (*c*) The environmental factor(s) required for development of BD as associated with the *HLA-A*26* antigen are unevenly distributed throughout the planet. (*d*) In these populations other genetic and environmental factors are of major importance in the development of BD; outpacing largely the association between *HLA-A*26* and BD.

In conclusion, we have performed the first GWAS in BD and provided a picture of the genetic constitution of BD susceptibility. It clearly, and perhaps somewhat unexpectedly, appears that the bulk of BD genetics remains in the MHC, where through extensive genotyping and haplotype analyses we were able to identify the *HLA-A* locus, as a second, independent, HLA-based

Table 3 Allele and phenotype frequencies of *HLA-A*26* subtypes in all subjects and in *HLA-B*5101* non-carriers

Allele/ phenotype	All subjects					*HLA-B*5101* non-carriers				
	N (frequency)					N (frequency)				
	BD	Controls	OR (95% CI)	p	Pc	BD	Controls	OR (95% CI)	p	Pc
Allele	n=600	n=600				n=272	n=510			
*A*2601*	79 (0.132)	44 (0.073)	1.92 (1.31 to 2.83)	0.00084	0.016	48 (0.178)	39 (0.076)	2.61 (1.66 to 4.10)	2.84×10^{-5}	0.00051
*A*2602*	18 (0.030)	9 (0.015)				13 (0.048)	7 (0.014)	3.63 (1.43 to 9.22)	0.0051	NS
*A*2603*	21 (0.035)	12 (0.020)				9 (0.033)	8 (0.016)			
*A*2605*	2 (0.003)	0 (0.000)				0 (0.000)	0 (0.0)			
*A*26*	120 (0.201)	65 (0.108)	2.07 (1.49 to 2.86)	9.99×10^{-6}	7.99×10^{-5}	70 (0.259)	54 (0.106)	2.96 (2.00 to 4.37)	4.54×10^{-8}	3.63×10^{-7}
Phenotype	n=300	n=300				n=136	n=255			
*A*2601*	77 (0.257)	42 (0.140)	2.12 (1.40 to 3.22)	0.00047	0.0089	46 (0.338)	37 (0.145)	3.01 (1.83 to 4.95)	1.21×10^{-5}	0.00022
*A*2602*	18 (0.060)	9 (0.030)				13 (0.096)	7 (0.027)	3.74 (1.46 to 9.62)	0.0048	NS
*A*2603*	21 (0.070)	12 (0.040)				9 (0.066)	8 (0.031)			
*A*2605*	2 (0.007)	0 (0.000)				0 (0.000)	0 (0.000)			
*A*26*	114 (0.380)	59 (0.197)	2.50 (1.73 to 3.62)	9.76×10^{-7}	7.81×10^{-6}	66 (0.485)	50 (0.196)	3.86 (2.45 to 6.10)	4.00×10^{-9}	3.20×10^{-8}

BD, Behçet disease; NS, not significant; OR, odds ratio; Pc, corrected p value.

susceptibility locus for BD. Functional studies could now be planned to unravel the joint action of two HLA alleles in BD pathogenesis.

Acknowledgements We sincerely thank the patients with BD who participated in this study. We also thank R Nishizaki for technical assistance.

Funding This work was supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan; a grant from the Ministry of Health, Labour and Welfare, Japan; and a grant from the Johnson & Johnson K.K. Vision Care Company. SB's laboratory is supported by the 'Agence Nationale pour la Recherche' (ANR), the 'Fédération des Maladies Orphelines' and the 'Association française de la maladie de Behçet'.

Competing interests None.

Ethics approval This study was conducted with the approval of the respective ethics authorities within each recruiting institution-that is, Yokohama City University, Hokkaido University, Kurume University, Yuasa Eye Clinic and Fujioka Eye Hospital; all in Japan.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Kaklamani VG**, Vaiopoulos G, Kaklamanis PG. Behçet's Disease. *Semin Arthritis Rheum* 1998;**27**:197–217.
2. **Ohno S**, Ohguchi M, Hirose S, *et al.* Close association of HLA-Bw51 with Behçet's disease. *Arch Ophthalmol* 1982;**100**:1455–8.
3. **Yabuki K**, Mizuki N, Ota M, *et al.* Association of MICA gene and HLA-B*5101 with Behçet's disease in Greece. *Invest Ophthalmol Vis Sci* 1999;**40**:1921–6.
4. **Mizuki N**, Yabuki K, Ota M, *et al.* Analysis of microsatellite polymorphism around the HLA-B locus in Iranian patients with Behçet's disease. *Tissue Antigens* 2002;**60**:396–9.
5. **Mizuki N**, Ota M, Katsuyama Y, *et al.* Association analysis between the MIC-A and HLA-B alleles in Japanese patients with Behçet's disease. *Arthritis Rheum* 1999;**42**:1961–6.
6. **Ota M**, Mizuki N, Katsuyama Y, *et al.* 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* 1999;**64**:1406–10.
7. **Mizuki N**, Ota M, Yabuki K, *et al.* Localization of the pathogenic gene of Behçet's disease by microsatellite analysis of three different populations. *Invest Ophthalmol Vis Sci* 2000;**41**:3702–8.
8. **Verity DH**, Wallace GR, Vaughan RW, *et al.* Behçet's disease: from Hippocrates to the third millennium. *Br J Ophthalmol* 2003;**87**:1175–83.
9. **Salvarani C**, Boiardi L, Casali B, *et al.* Endothelial nitric oxide synthase gene polymorphisms in Behçet's disease. *J Rheumatol* 2002;**29**:535–40.
10. **Karasneh JA**, Hajeer AH, Silman A, *et al.* Polymorphisms in the endothelial nitric oxide synthase gene are associated with Behçet's disease. *Rheumatology (Oxford)* 2005;**44**:614–17.
11. **Karasneh J**, Gül A, Ollier WE, *et al.* Whole-genome screening for susceptibility genes in multicase families with Behçet's disease. *Arthritis Rheum* 2005;**52**:1836–42.
12. **Risch NJ**. Searching for genetic determinants in the new millennium. *Nature* 2000;**405**:847–56.
13. **International Study Group for Behçet's Disease**. Criteria for the diagnosis of Behçet's disease. *Lancet* 1990;**335**:1078–80.
14. **Collins HE**, Li H, Inda SE, *et al.* A simple and accurate method for determination of microsatellite total allele content differences between DNA pools. *Hum Genet* 2000;**106**:218–26.
15. **Oka A**, Hayashi H, Tomizawa M, *et al.* Localization of a non-melanoma skin cancer susceptibility region within the major histocompatibility complex by association analysis using microsatellite markers. *Tissue Antigens* 2003;**61**:203–10.
16. **Tamiya G**, Shinya M, Imanishi T, *et al.* Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. *Hum Mol Genet* 2005;**14**:2305–21.
17. **Itoh Y**, Mizuki N, Shimada T, *et al.* High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics* 2005;**57**:717–29.
18. **Matte C**, Lacaille J, Zijenah L, *et al.* HLA-G and HLA-E polymorphisms in an indigenous African population. The ZVITAMBO Study Group. *Hum Immunol* 2000;**61**:1150–6.
19. **Martinez-Laso J**, Gomez-Casado E, Arnaiz-Villena A. HLA-E and HLA-G typing. *Methods Mol Biol* 2003;**210**:223–36.
20. **Hviid TV**, Hylenius S, Hoegh AM, *et al.* HLA-G polymorphisms in couples with recurrent spontaneous abortions. *Tissue Antigens* 2002;**60**:122–32.
21. **Itoh Y**, Inoko H, Kulski JK, *et al.* Four-digit allele genotyping of the HLA-A and HLA-B genes in Japanese patients with Behcet's disease by a PCR-SSOP-Luminex method. *Tissue Antigens* 2006;**67**:390–4.
22. **Chung YM**, Yeh TS, Sheu MM, *et al.* [Behcet's disease with ocular involvement in Taiwan: a joint survey of six major ophthalmological departments]. *J Formos Med Assoc* 1990;**89**:413–17.
23. **Mizuki N**, Ohno S, Ando H, *et al.* A strong association between HLA-B*5101 and Behçet's disease in Greek patients. *Tissue Antigens* 1997;**50**:57–60.
24. **Park KS**, Park JS, Nam JH, *et al.* HLA-E*0101 and HLA-G*010101 reduce the risk of Behcet's disease. *Tissue Antigens* 2007;**69**:139–44.
25. **Fischer G**, Wittmann-Liebold B, Lang K, *et al.* Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* 1989;**337**:476–8.
26. **Freskgård PO**, Bergenhem N, Jonsson BH, *et al.* Isomerase and chaperone activity of prolyl isomerase in the folding of carbonic anhydrase. *Science* 1992;**258**:466–8.
27. **Handschumacher RE**, Harding MW, Rice J, *et al.* Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science* 1984;**226**:544–7.
28. **Sherry B**, Zybarth G, Alfano M, *et al.* Role of cyclophilin A in the uptake of HIV-1 by macrophages and T lymphocytes. *Proc Natl Acad Sci USA* 1998;**95**:1758–63.
29. **Montague JW**, Gaido ML, Frye C, *et al.* A calcium-dependent nuclease from apoptotic rat thymocytes is homologous with cyclophilin. Recombinant cyclophilins A, B, and C have nuclease activity. *J Biol Chem* 1994;**269**:18877–80.

30. **Foster JW**, Dominguez-Steglich MA, Guioli S, *et al.* Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 1994;**372**:525–30.
31. **Southard-Smith EM**, Kos L, Pavan WJ. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat Genet* 1998;**18**:60–4.
32. **Hagiwara N**, Klewer SE, Samson RA, *et al.* Sox6 is a candidate gene for p100H myopathy, heart block, and sudden neonatal death. *Proc Natl Acad Sci USA* 2000;**97**:4180–5.
33. **Kamachi Y**, Uchikawa M, Collignon J, *et al.* Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. *Development* 1998;**125**:2521–32.
34. **van de Wetering M**, Oosterwegel M, van Norren K, *et al.* Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J* 1993;**12**:3847–54.
35. **Schilham MW**, Oosterwegel MA, Moerer P, *et al.* Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature* 1996;**380**:711–14.
36. **Kidd T**, Brose K, Mitchell KJ, *et al.* Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell* 1998;**92**:205–15.
37. **Padlan EA**. Anatomy of the antibody molecule. *Mol Immunol* 1994;**31**:169–217.
38. **Sidebottom D**, Grennan DM, Sanders PA, *et al.* Immunoglobulin lambda light chain genes in rheumatoid arthritis. *Ann Rheum Dis* 1987;**46**:587–9.
39. **Queiroz RG**, Tamia-Ferreira MC, Carvalho IF, *et al.* Association between EcoRI fragment-length polymorphism of the immunoglobulin lambda variable 8 (IGLV8) gene family with rheumatoid arthritis and systemic lupus erythematosus. *Braz J Med Biol Res* 2001;**34**:525–8.
40. **Kikkawa T**, Shirotsuki H, Ueno H. Serum immunoglobulin levels in various kinds of uveitis such as Behçet's disease, Vogt Koyanagi Harada's disease, and chronic recurrent monocular and binocular panuveitis. *Folia Ophthal Jap* 1975;**26**:315–20.
41. **Moriyama H**, Matsumoto K, Mimura Y. [Serum immunoglobulin D in Harada disease and Behçet's disease (author's transl)]. *Nippon Ganka Gakkai Zasshi* 1976;**80**:480–5.
42. **Scully C**, Boyle P, Yap PL. Immunoglobulins G, M, A, D and E in Behçet's syndrome. *Clin Chim Acta* 1982;**120**:237–42.
43. **Cengiz K**. Serum IgE concentrations in complete Behçet's disease. *J Clin Pathol* 1990;**43**:262.
44. **Shiina T**, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens* 2004;**64**:631–49.
45. **Yabuki K**, Ohno S, Mizuki N, *et al.* HLA class I and II typing of the patients with Behçet's disease in Saudi Arabia. *Tissue Antigens* 1999;**54**:273–7.
46. **O'Brien SJ**, Gao X, Carrington M. HLA and AIDS: a cautionary tale. *Trends Mol Med* 2001;**7**:379–81.
47. **Creemers PC**, Kahn D, Hill JC. HLA-A and -B alleles in cornea donors as risk factors for graft rejection. *Transpl Immunol* 1999;**7**:15–18.
48. **Weisdorf D**, Hakke R, Blazar B, *et al.* Risk factors for acute graft-versus-host disease in histocompatible donor bone marrow transplantation. *Transplantation* 1991;**51**:1197–203.
49. **Mohyeddin Bonab M**, Alimoghaddam K, Vatandoust S, *et al.* Are HLA antigens a risk factor for acute GVHD in thalassemic patients receiving HLA-identical stem cell transplantation? *Transplant Proc* 2004;**36**:3190–3.
50. **Celik G**, Sen ES, Ulger AF, *et al.* Human leukocyte antigens A and B in Turkish patients with sarcoidosis. *Arch Bronconeumol* 2004;**40**:449–52.
51. **Suzuki T**, Mitsuishi Y, Sano Y, *et al.* Phlyctenular keratitis associated with meibomitis in young patients. *Am J Ophthalmol* 2005;**140**:77–82.
52. **Verity DH**, Wallace GR, Vaughan RW, *et al.* HLA and tumour necrosis factor (TNF) polymorphisms in ocular Behçet's disease. *Tissue Antigens* 1999;**54**:264–72.
53. **Mizuki N**, Ota M, Katsuyama Y, *et al.* HLA class I genotyping including HLA-B*51 allele typing in the Iranian patients with Behçet's disease. *Tissue Antigens* 2001;**57**:457–62.
54. **Kilmartin DJ**, Finch A, Acheson RW. Primary association of HLA-B51 with Behçet's disease in Ireland. *Br J Ophthalmol* 1997;**81**:649–53.
55. **Kera J**, Mizuki N, Ota M, *et al.* Significant associations of HLA-B*5101 and B*5108, and lack of association of class II alleles with Behçet's disease in Italian patients. *Tissue Antigens* 1999;**54**:565–71.
56. **Pirim I**, Atasoy M, Ikbal M, *et al.* HLA class I and class II genotyping in patients with Behcet's disease: a regional study of eastern part of Turkey. *Tissue Antigens* 2004;**64**:293–7.

Molecular Vision 2009; 15:1115-1119 <http://www.molvis.org/molvis/v15/a118>
Received 15 April 2009 | Accepted 29 May 2009 | Published 3 June 2009

Evaluation of *PTPN22* polymorphisms and Vogt-Koyanagi-Harada disease in Japanese patients

Yukihiro Horie,[1] Nobuyoshi Kitaichi,[1] Yoshihiko Katsuyama,[2] Kazuhiko Yoshida,[1] Toshie Miura,[1] Masao Ota,[3] Yuri Asukata,[4] Hidetoshi Inoko,[5] Nobuhisa Mizuki,[4] Susumu Ishida,[1] Shigeaki Ohno[1,6]

[1]Department of Ophthalmology, Hokkaido University Graduate School of Medicine, Sapporo, Japan; [2]Department of Pharmacy, Shinshu University School of Medicine, Matsumoto, Japan; [3]Department of Legal Medicine, Shinshu University School of Medicine, Matsumoto, Japan; [4]Department of Ophthalmology, Yokohama City University School of Medicine, Yokohama, Japan; [5]Department of Molecular Life Science, Division of Molecular Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Japan; [6]Department of Ocular Inflammation and Immunology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

Purpose: Vogt-Koyanagi-Harada (VKH) disease is an autoimmune disorder against melanocytes. Polymorphisms of the protein tyrosine phosphatase non-receptor 22 gene (*PTPN22*) have recently been reported to be associated with susceptibility to several autoimmune diseases. In this study, genetic susceptibility to VKH disease was investigated by screening for single nucleotide polymorphisms (SNPs) of *PTPN22*.
Methods: A total of 167 Japanese patients with VKH disease and 188 healthy Japanese controls were genotyped by direct sequencing methods for six SNPs (rs3811021, rs1217413, rs1237682, rs3761935, rs3789608, and rs2243471) of *PTPN22* including the uncoding exons.
Results: The six SNPs in *PTPN22* showed no significant association with susceptibility to VKH disease or its ocular, neurologic, or dermatological manifestation.
Conclusions: Further studies are needed to clarify the genetic mechanisms underlying VKH disease.

Vogt-Koyanagi-Harada (VKH) disease is one of the most frequent forms of uveitis in Japan [1]. It is characterized as bilateral panuveitis accompanied by neurologic and skin lesions [2,3]. This disease is considered to be an autoimmune disease against melanocytes [4,5]. Though the etiology of VKH disease still remains unknown, genetic factors may play an important role in susceptibility as indicated by an established association between VKH disease and specific human leukocyte antigen (*HLA*)-*DRB1* alleles [6,7].

The protein tyrosine phosphatase non-receptor 22 gene (*PTPN22*) is located on chromosome 1p13.3-p13.1, and it encodes the lymphoid-specific phosphatase (Lyp) that is important in the negative control of T-cell activation and development [8-10]. Recently, it was reported the single nucleotide polymorphism (SNP), R620W (rs2476601), in *PTPN22* increased susceptibility to several autoimmune diseases including rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and insulin dependent diabetes mellitus (IDDM) [11-15]. In the *PTPN22* risk variant (rs2476601), this substitution disrupts an interaction between Lyp and the protein tyrosine kinase, Csk, and may translate biologically to a potential for 'hyperreactive' pathogenic T-cell responses [8].

This R620W mutation was not observed in the Japanese population [16,17]. Therefore, in this study, we analyzed six SNPs, which belong to the same haplotype block as R620W (rs2476601) in *PTPN22*. For the efficacy of the linkage analysis, we chose six SNPs of which minor allele frequencies were more than 15% from the database of Japanese Single Nucleotide Polymorphisms [18,19].

METHODS

We recruited 167 VKH (72 males and 95 females) patients and 188 healthy controls for this study. All patients and control subjects were Japanese. Patients were diagnosed according to the "Revised Diagnostic Criteria for VKH Disease" [3] at the Uveitis Survey Clinic of the Hokkaido University Hospital (Sapporo, Japan) and Yokohama City University Hospital (Yokohama, Japan). All patients and control subjects were informed of the study's purpose, and their consent obtained. The study was approved by the ethics committee from each institute participating in this study.

DNA was prepared from peripheral blood specimens using the QIAamp DNA Blood Mini Kit (Qiagen, Tokyo, Japan). Six SNPs (rs3811021, rs1217413, rs1237682, rs3761935, rs3789608, and rs2243471) from the *PTPN22* region were examined (Figure 1). Each of the six SNPs was amplified by standard polymerase chain reactions (PCRs; Table 1). After purification using ExoSAP-IT (USB

Correspondence to: Yukihiro Horie, MD, Department of Ophthalmology and Visual Sciences, Hokkaido University Graduate School of Medicine, N-15, W-7, Kita-ku, Sapporo, 060-8638, Japan; Phone: +81-11-706-5944; FAX: +81-11-706-5948; email: y-horie@med.hokudai.ac.jp

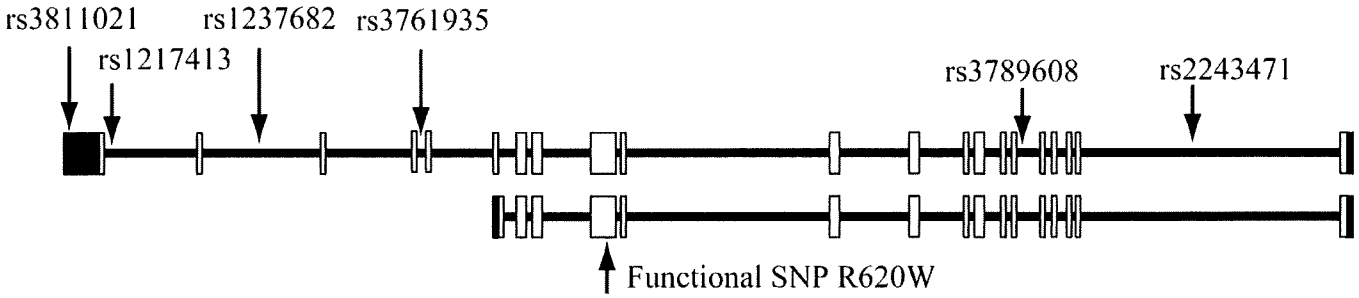

Figure 1. *PTPN22* structure with two transcript isoforms and six SNP. Six SNP variants with minor allele frequencies 15% from the database of Japanese Single Nucleotide Polymorphisms. The black and white areas in the exons indicate the UTR and coding region, respectively.

Table 1. PCR primers for *PTPN22* SNPs.

SNP		position	Primers	Product size (bp)	sequence primer
rs3811021	(SNP1)	114158186	F: TGGGTTGCAATACAAACTGCTC R: TCAATTTGCCCTATTGGACTTC	600	Forward
rs1217413	(SNP2)	114159273	F: TTGCAGGTGTACTTGCAGCC R: TTGAAGGATTTCTGGACCGAC	552	Forward
rs1237682	(SNP3)	114165627	F: AAGGAGGCACAGATTCCACAC R: TGACCATGCCAATATACCAACTG	589	Forward
rs3761935	(SNP4)	114174051	F: AAAGTTTCCGGCATGTTTCC R: TGGTGATTGTCGGCTAAGATTG	595	Reverse
rs3789608	(SNP5)	114199311	F: CATCATGGTCTGGCCAATTC R: TGAGGTGGAGTTCTAACCACAAG	589	Forward
rs2243471	(SNP6)	114207525	F: GACAAGACTGAATTGTACGAGCG R: CACCATCTCCAGCCTCTCAC	577	Forward

The position of the SNPs is cited from the NCBI database.

Corporation, Cleveland, OH), the PCR products were sequenced with Big Dye Terminator v3.1 (Applied Biosystems, Foster City, CA) using either sense or antisense primers (Table 1). The BigDye XTerminator Purification Kit (Applied Biosystems) was used to purify the DNA from sequencing reactions. The sequencing reactions were analyzed using an ABI3130 sequencer (Applied Biosystems).

Statistical analysis: For statistical analyses, the Hardy–Weinberg equilibrium was tested for each SNP among the control subjects. Genotype frequency differences between the case and control genotypes were assessed by the χ^2 test. The calculation of linkage disequilibrium (LD) and pair-wise LD (D' value) between SNPs of the *PTPN22* region and the haplotypes was performed with Haploview software, version 3.32. The maximum likelihood estimates of haplotype frequencies were estimated by pairs of unphased genotypes using the expectation-maximization (EM) algorithms in the R package 'haplo.stats' [20].

RESULTS

Allele frequencies for the six SNPs covering the gene were in Hardy–Weinberg equilibrium in both the patients and controls. The allelic frequency of each SNP in both groups was nearly equal, and no association was detected when compared independently (odds ratio, OR 1.14–1.35; Table 2). Stratifying the patients by the presence of diffuse choroiditis, sunset glow fundus, nummular chorioretinal depigmented spots, neurologic auditory involvement, meningismus, tinnitus, cerebrospinal fluid pleocytosis, or integumentary findings also revealed no evidence of association in VKH disease (data not shown). We calculated pairwise D' values for all SNP pairs in *PTPN22* (Figure 2). The pairwise D' values in the gene were nearly 1 among almost all SNP pairs, indicating the SNPs were highly associated with each other and the entire *PTPN22* was contained within a single LD block. Haplotype analysis predicted and revealed that *PTPN22* was not associated with VKH disease in this Japanese cohort (data not shown).

DISCUSSION

In the present study, we analyzed polymorphisms of the new candidate gene, *PTPN22*, in Japanese patients with VKH

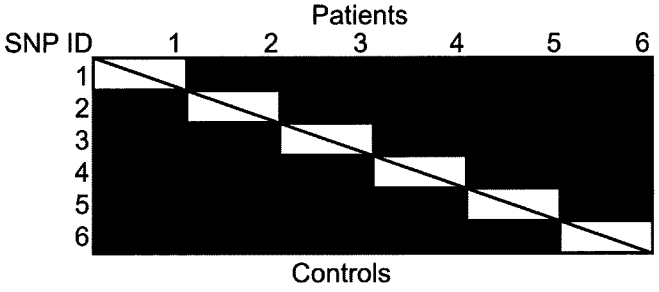

Figure 2. D' score for the six SNPs studied across the *PTPN22* haplotype. Black cells indicate that D' is greater than 0.9. Upper: patient population, lower: control population. The figure indicates that the six SNPs were in all the same haplotype block.

Table 2. Genotype frequencies in VKH patients and controls.

SNP	Allele	VKH (n=168)	Percentage	Control (n=187)	Percentage	Odds ratio (95% CI)	p
rs3811021	C/C	2	1.2	9	4.8	0.24 (0.05–17.47)	0.05
	T/C	56	33.5	67	35.8	0.90 (0.58–7.38)	0.65
	T/T	109	65.3	111	59.4	1.29 (0.84–7.53)	0.25
	C	60		85		0.75 (0.51–7.41)	0.12
rs1217413	A/A	26	16	36	19.4	0.79 (0.45–7.63)	0.41
	A/G	77	47.2	91	48.9	0.93 (0.61–7.43)	0.75
	G/G	60	36.8	59	31.7	1.25 (0.80–7.56)	0.32
	A	129		163		0.84 (0.62–7.25)	0.26
rs1237682	C/C	26	15.5	33	18.4	0.83 (0.47–7.63)	0.53
	T/C	79	47	89	49.7	0.94 (0.62–7.34)	0.77
	T/T	59	35.1	57	31.8	1.20 (0.77–7.53)	0.42
	C	131		155		0.87 (0.64–7.24)	0.37
rs3761935	G/G	3	1.8	10	5.3	0.33 (0.09–12.76)	0.08
	T/G	55	33.3	67	35.8	0.90 (0.58–7.39)	0.62
	T/T	107	64.8	110	58.8	1.29 (0.84–7.54)	0.25
	G	61		87		0.75 (0.52–7.41)	0.12
rs3789608	T/T	2	1.2	9	4.8	0.24 (0.05–17.57)	0.05
	C/T	57	34.1	68	36.4	0.91 (0.59–7.38)	0.66
	C/C	108	64.7	110	58.8	1.26 (0.82–7.50)	0.29
	T	61		86		0.74 (0.51–7.41)	0.11
rs2243471	A/A	27	16.7	37	20.2	0.79 (0.46–7.62)	0.4
	A/G	79	48.8	88	48.1	1.03 (0.67–7.36)	0.9
	G/G	56	34.6	58	31.7	1.14 (0.73–7.48)	0.57
	A	133		162		0.88 (0.65–7.24)	0.39

The above table is the genotype and allele frequencies of the VKH patients and healthy controls. There are no differences between patients and controls.

disease. The gene encodes an important negative regulator of T cell activation [9]. An SNP of *PTPN22*, R620W (rs2476601) was reported to be associated with several autoimmune diseases such as RA, SLE, and IDDM [11,12, 14,15]. However, this SNP, which disrupts an interaction between Lyp and the protein tyrosine kinase, Csk, does not exist as a polymorphism in the Japanese population [9,10, 12]. Therefore, in this study, we examined six other SNPs to evaluate the susceptibility locus of *PTPN22*. *HLA-DRB1* is a common genetic factor in autoimmune diseases (RA and IDDM). Therefore, there may be other common genetic factors in VKH disease [21].

VKH disease is considered to be an autoimmune disease against melanocytes [2-5]. In early studies, activated T lymphocytes were elevated and attacked melanocytes of ocular choroidal tissue in patients in the active phase of VKH disease [22]. Antigen-specific T-cell assay revealed that peptide fragments of the tyrosinase family proteins (tyrosinase, tyrosinase related protein 1 and 2) proliferated in T lymphocytes collected from VKH patients [4,5]. These proteins are found in human melanocytes. These antigen-specific T cell responses were detected in cells collected from *HLA-DRB1*04* positive VKH patients only but not from *HLA-DRB1*04* negative patients or *HLA-DRB1*04* positive healthy people [7,23]. In the Japanese population, 40% of healthy people have *HLA-DRB1*04* [7]. However, people having VKH disease represent only 0.01% of the Japanese population [1,7,24,25]. In addition, some patients with VKH disease are *HLA-DRB1*04* negative [7]. Thus, it is believed *HLA-DRB1*04* is a major susceptible gene in VKH disease. However, other minor genetic factors still remain unclear. To find other susceptible genes, we studied the tyrosinase gene (*TYR*), tyrosinase related protein 1 gene (*TYRP1*), tyrosinase related protein 2 gene (*TYRP2*), and interferon (*IFN-γ*), but we could not find any association with these genes and VKH disease [7,26]. Genetic influences of VKH were also investigated in other countries, but the etiology of the disease seems to be unresolved [27-29].

In this study, we found no association between *PTPN22* and VKH disease in the individuals studied. Our results suggest that further molecular genetic studies are needed to detect novel genetic loci and predisposing genes and to elucidate the true genetic mechanisms underlying VKH disease.

ACKNOWLEDGMENTS

This work was supported by Grant-in-Aid for Scientific Research (Category B, No. 19406028) from the Ministry of Education, Culture, Science, Sports and Technology (MEXT), Japan.

REFERENCES

1. Goto H, Mochizuki M, Yamaki K, Kotake S, Usui M, Ohno S. Epidemiological survey of intraocular inflammation in Japan. Jpn J Ophthalmol 2007; 51:41-4. [PMID: 17295139]
2. Kitamura M, Takami K, Kitaichi N, Namba K, Kitamei H, Kotake S, Ohno S. Comparative study of two sets of criteria for the diagnosis of Vogt-Koyanagi-Harada's disease. Am J Ophthalmol 2005; 139:1080-5. [PMID: 15953440]

3. Read RW, Holland GN, Rao NA, Tabbara KF, Ohno S, Arellanes-Garcia L, Pivetti-Pezzi P, Tessler HH, Usui M. Revised diagnostic criteria for Vogt-Koyanagi-Harada disease: report of an international committee on nomenclature. Am J Ophthalmol 2001; 131:647-52. [PMID: 11336942]
4. Yamaki K, Gocho K, Hayakawa K, Kondo I, Sakuragi S. Tyrosinase family proteins are antigens specific to Vogt-Koyanagi-Harada disease. J Immunol 2000; 165:7323-9. [PMID: 11120868]
5. Yamaki K, Takiyama N, Itho N, Mizuki N, Seiya M, Sinsuke W, Hayakawa K, Kotani T. Experimentally induced Vogt-Koyanagi-Harada disease in two Akita dogs. Exp Eye Res 2005; 80:273-80. [PMID: 15670805]
6. Islam SM, Numaga J, Fujino Y, Hirata R, Matsuki K, Maeda H, Masuda K. HLA class II genes in Vogt-Koyanagi-Harada disease. Invest Ophthalmol Vis Sci 1994; 35:3890-6. [PMID: 7928186]
7. Horie Y, Takemoto Y, Miyazaki A, Namba K, Kase S, Yoshida K, Ota M, Hasumi Y, Inoko H, Mizuki N, Ohno S. Tyrosinase gene family and Vogt-Koyanagi-Harada disease in Japanese patients. Mol Vis 2006; 12:1601-5. [PMID: 17200659]
8. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004; 36:337-8. [PMID: 15004560]
9. Siminovitch KA. PTPN22 and autoimmune disease. Nat Genet 2004; 36:1248-9. [PMID: 15565104]
10. Gregersen PK. Gaining insight into PTPN22 and autoimmunity. Nat Genet 2005; 37:1300-2. [PMID: 16314859]
11. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM, Conn MT, Chang M, Chang SY, Saiki RK, Catanese JJ, Leong DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batliwalla F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Amos CI, Sninsky JJ, Gregersen PK. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004; 75:330-7. [PMID: 15208781]
12. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, Chang M, Ramos P, Baechler EC, Batliwalla FM, Novitzke J, Williams AH, Gillett C, Rodine P, Graham RR, Ardlie KG, Gaffney PM, Moser KL, Petri M, Begovich AB, Gregersen PK, Behrens TW. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am J Hum Genet 2004; 75:504-7. [PMID: 15273934]
13. Michou L, Lasbleiz S, Rat AC, Migliorini P, Balsa A, Westhovens R, Barrera P, Alves H, Pierlot C, Glikmans E, Garnier S, Dausset J, Vaz C, Fernandes M, Petit-Teixeira E, Lemaire I, Pascual-Salcedo D, Bombardieri S, Dequeker J, Radstake TR, Van Riel P, van de Putte L, Lopes-Vaz A, Prum B, Bardin T, Dieude P, Cornelis F. Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene. Proc Natl Acad Sci USA 2007; 104:1649-54. [PMID: 17237219]
14. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W, Lee AT, Ortmann W, Behrens TW, Gregersen PK. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 2005; 76:561-71. [PMID: 15719322]
15. Wu H, Cantor RM, Graham DS, Lingren CM, Farwell L, Jager PL, Bottini N, Grossman JM, Wallace DJ, Hahn BH, Julkunen H, Hebert LA, Rovin BH, Birmingham DJ, Rioux JD, Yu CY, Kere J, Vyse TJ, Tsao BP. Association analysis of the R620W polymorphism of protein tyrosine phosphatase PTPN22 in systemic lupus erythematosus families: increased T allele frequency in systemic lupus erythematosus patients with autoimmune thyroid disease. Arthritis Rheum 2005; 52:2396-402. [PMID: 16052563]
16. Ichimura M, Kaku H, Fukutani T, Koga H, Mukai T, Miyake I, Yamada K, Koda Y, Hiromatsu Y. Associations of protein tyrosine phosphatase nonreceptor 22 (PTPN22) gene polymorphisms with susceptibility to Graves' disease in a Japanese population. Thyroid 2008; 18:625-30. [PMID: 18578611]
17. Ikari K, Momohara S, Inoue E, Tomatsu T, Hara M, Yamanaka H, Kamatani N. Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. Rheumatology (Oxford) 2006; 45:1345-8. [PMID: 16690758]
18. Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res 2002; 30:158-62. [PMID: 11752280]
19. Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T. Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. J Hum Genet 2002; 47:605-10. [PMID: 12436197]
20. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 2002; 70:425-34. [PMID: 11791212]
21. Stastny P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. N Engl J Med 1978; 298:869-71. [PMID: 147420]
22. Moorthy RS, Inomata H, Rao NA. Vogt-Koyanagi-Harada syndrome. Surv Ophthalmol 1995; 39:265-92. [PMID: 7725227]
23. Shindo Y, Inoko H, Yamamoto T, Ohno S. HLA-DRB1 typing of Vogt-Koyanagi-Harada's disease by PCR-RFLP and the strong association with DRB1*0405 and DRB1*0410. Br J Ophthalmol 1994; 78:223-6. [PMID: 7908535]
24. Ohno S, Nakamura S, Hori S, Shimakawa M, Kawashima H, Mochizuki M, Sugita S, Ueno S, Yoshizaki K, Inaba G. Efficacy, safety, and pharmacokinetics of multiple administration of infliximab in Behcet's disease with refractory uveoretinitis. J Rheumatol 2004; 31:1362-8. [PMID: 15229958]
25. Saito S, Ota S, Yamada E, Inoko H, Ota M. Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population. Tissue Antigens 2000; 56:522-9. [PMID: 11169242]

26. Horie Y, Kitaichi N, Takemoto Y, Namba K, Yoshida K, Hirose S, Hasumi Y, Ota M, Inoko H, Mizuki N, Ohno S. Polymorphism of IFN-gamma gene and Vogt-Koyanagi-Harada disease. Mol Vis 2007; 13:2334-8. [PMID: 18199975]
27. Du L, Yang P, Hou S, Lin X, Zhou H, Huang X, Wang L, Kijlstra A. Association of the CTLA-4 gene with Vogt-Koyanagi-Harada syndrome. Clin Immunol 2008; 127:43-8. [PMID: 18282809]
28. Hou S, Yang P, Du L, Zhou H, Lin X, Liu X, Kijlstra A. Small ubiquitin-like modifier 4 (SUMO4) polymorphisms and Vogt-Koyanagi-Harada (VKH) syndrome in the Chinese Han population. Mol Vis 2008; 14:2597-603. [PMID: 19122825]
29. Meng Q, Liu X, Yang P, Hou S, Du L, Zhou H, Kijlstra A. PDCD1 genes may protect against extraocular manifestations in Chinese Han patients with Vogt-Koyanagi-Harada syndrome. Mol Vis 2009; 15:386-92. [PMID: 19234630]

The print version of this article was created on 29 May 2009. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.

Gene Expression Profile for Predicting Survival in Advanced-Stage Serous Ovarian Cancer Across Two Independent Datasets

Kosuke Yoshihara[1], Atsushi Tajima[2], Tetsuro Yahata[1], Shoji Kodama[3], Hiroyuki Fujiwara[4], Mitsuaki Suzuki[4], Yoshitaka Onishi[5], Masayuki Hatae[5], Kazunobu Sueyoshi[6], Hisaya Fujiwara[7], Yoshiki Kudo[7], Kohei Kotera[8], Hideaki Masuzaki[9], Hironori Tashiro[10], Hidetaka Katabuchi[10], Ituro Inoue[2], Kenichi Tanaka[1]*

1 Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan, **2** Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan, **3** Department of Gynecology, Niigata Cancer Center Hospital, Niigata, Japan, **4** Department of Obstetrics and Gynecology, Jichi Medical University, Shimotsuke, Japan, **5** Department of Obstetrics and Gynecology, Kagoshima City Hospital, Kagoshima, Japan, **6** Department of Pathology, Kagoshima City Hospital, Kagoshima, Japan, **7** Department of Obstetrics and Gynecology, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan, **8** Department of Obstetrics and Gynecology, Nagasaki Municipal Hospital, Nagasaki, Japan, **9** Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan, **10** Department of Gynecology, Faculty of Medical and Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan

Abstract

Background: Advanced-stage ovarian cancer patients are generally treated with platinum/taxane-based chemotherapy after primary debulking surgery. However, there is a wide range of outcomes for individual patients. Therefore, the clinicopathological factors alone are insufficient for predicting prognosis. Our aim is to identify a progression-free survival (PFS)-related molecular profile for predicting survival of patients with advanced-stage serous ovarian cancer.

Methodology/Principal Findings: Advanced-stage serous ovarian cancer tissues from 110 Japanese patients who underwent primary surgery and platinum/taxane-based chemotherapy were profiled using oligonucleotide microarrays. We selected 88 PFS-related genes by a univariate Cox model ($p<0.01$) and generated the prognostic index based on 88 PFS-related genes after adjustment of regression coefficients of the respective genes by ridge regression Cox model using 10-fold cross-validation. The prognostic index was independently associated with PFS time compared to other clinical factors in multivariate analysis [hazard ratio (HR), 3.72; 95% confidence interval (CI), 2.66–5.43; $p<0.0001$]. In an external dataset, multivariate analysis revealed that this prognostic index was significantly correlated with PFS time (HR, 1.54; 95% CI, 1.20–1.98; $p=0.0008$). Furthermore, the correlation between the prognostic index and overall survival time was confirmed in the two independent external datasets (log rank test, $p=0.0010$ and 0.0008).

Conclusions/Significance: The prognostic ability of our index based on the 88-gene expression profile in ridge regression Cox hazard model was shown to be independent of other clinical factors in predicting cancer prognosis across two distinct datasets. Further study will be necessary to improve predictive accuracy of the prognostic index toward clinical application for evaluation of the risk of recurrence in patients with advanced-stage serous ovarian cancer.

Citation: Yoshihara K, Tajima A, Yahata T, Kodama S, Fujiwara H, et al. (2010) Gene Expression Profile for Predicting Survival in Advanced-Stage Serous Ovarian Cancer Across Two Independent Datasets. PLoS ONE 5(3): e9615. doi:10.1371/journal.pone.0009615

Editor: Zoltán Bochdanovits, VU University Medical Center and Center for Neurogenomics and Cognitive Research, The Netherlands

Received November 3, 2009; **Accepted** February 16, 2010; **Published** March 12, 2010

Funding: This work was supported in part by a Grant-in-Aid for the Third-term Cancer Control Strategy Program from the Ministry of Health, Labor and Welfare, Japan (KT), and 2009 Research and Study Program of Tokai University Educational System General Research Organization (AT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tanaken@med.niigata-u.ac.jp

Introduction

Patients with advanced-stage ovarian cancer generally undergo primary debulking surgery followed by platinum/taxane-based chemotherapy. Although postoperative introduction of taxane drug has improved the 5-year survival rate for advanced-stage ovarian cancer, patients with this cancer have a 5-year survival rate of only 30% [1–3]. Clinicopathological characteristics, such as debulking status after primary surgery, are clinically considered important indicators of prognosis [4,5]. However, recurrence after optimal debulking surgery occurs in some patients, while disease-free status after incomplete surgery is maintained in others. In fact, it has been reported that 34% of patients treated with optimal surgery and platinum-taxane combination chemotherapy for advanced-stage ovarian cancer recur within 12 months [4]. Therefore, these clinicopathological factors alone are insufficient for predicting prognosis and elucidating the pathological mechanisms of disease progression or recurrence. Molecular biology

approaches can be used to identify new prognosis-related profiles leading to elucidation of pathological issues of advanced-stage serous ovarian cancer.

Microarray technology has been developing very rapidly, and it has become relatively easy to analyze the expression levels of thousands of genes within cancer cells. Although many studies have reported the associations of gene expression profiles with prognoses in cancer patients [6–10], a limited number of such profiles are used in clinical settings. Microarray technology is clinically applied for predicting prognosis in breast cancer patients. MammaPrint ™ (Agendia BV, Amsterdam, the Netherlands) has been already put to practical use for the purpose. Meanwhile, there are no microarray kits for clinical diagnosis and management in patients with ovarian cancer yet.

Three studies have recently reported gene expression profiles that predict overall survival (OS) in ovarian cancer patients using microarray techniques [11–13]. These studies use a relative large sample size (n>80) for establishing a survival-related profile in a discovery phase of the experiment and an external independent dataset as the validation set to solve the problem that the number of the genomic variables examined is much larger than that of subjects. Thus, research on the overall survival-related profiles in ovarian cancer patients has progressed, whereas there are no extensive studies based on multicenter validation of gene expression profiles for prediction of disease progression or recurrence in patients with ovarian cancer [14–15]. Prediction of the risk of recurrence in patients with advanced-stage ovarian cancer receiving standard treatments (primary surgery+platinum/taxane-based chemotherapy) is more important with respect to optimization of clinical management [16].

We have recently reported that there are high similarities in gene expression between early-stage and a subset of advanced-stage serous ovarian cancer patients that have favorable prognoses, and two molecular subgroups among patients with advanced-stage serous ovarian cancer according to gene expression profiles reflecting tumor progression and prognosis [17]. In this study, we focused on progression-free survival (PFS) time in a larger number of patients only with advanced-stage serous ovarian cancer treated with platinum/taxane-based chemotherapy, and tried to identify PFS-related gene expression profile using a new survival analysis method: ridge regression Cox model [18]. We then assessed the correlation between our PFS-related genes expression profile and survival time in an external independent dataset of advanced-stage serous ovarian cancer.

Results

Clinical Characteristics

The clinical characteristics of 110 Japanese patients with advanced-stage serous ovarian cancer are summarized in Table 1. In the discovery set, 93 patients (84.5%) were diagnosed as the International Federation of Gynecology and Obstetrics (FIGO) stage III, and 17 patients (15.5%) as FIGO stage IV [19]. All patients received platinum/taxane-based chemotherapy after primary surgery. The median progression-free and overall survival times were 17 and 31 months, respectively.

On the other hand, we used a part of publicly available microarray data (GSE9891) as an external independent dataset (See Materials and Methods) [20]. The clinical characteristics of 87 patients with advanced-stage serous ovarian cancer in the external dataset are listed in Table S1 [20]. Kaplan-Meier survival analysis showed that there were no significant differences in PFS and OS time between patients of the discovery dataset and those of the external dataset (Figure S1). When we compared clinicopathological characteristics between the discovery set and the external dataset, there were significant differences in frequencies of stage (Table S1). Because grading system adopted in the external dataset was distinct from that in the discovery set [21–23], we could not make a simple comparison of malignant grade between the two datasets. Then we examined the association between clinicopathological features and PFS time in patients with advanced-stage serous ovarian cancer of each dataset. Multivariate analysis revealed that only optimal surgery was an independent prognostic factor for PFS in the discovery dataset (Table S2) and that there was marginally significant correlation between debulking status of primary surgery and PFS time in the external dataset (Table S2). Therefore, we planned first to develop a prognostic index based on PFS-related genes in the discovery dataset, secondarily to evaluate the prognostic ability of our index in the external dataset using multivariate analysis, and then thirdly to assess predictive performance of the prognostic index again after the stratification of patients according to the debulking status of primary surgery.

Table 1. Clinical characteristics of advanced-stage serous ovarian cancer patients.

	Present Dataset (n = 110)	Percentage
Median age, years (range)	58 (23–85)	
Stage		
Stage III	93	84.5
Stage IV	17	15.5
CA125 (IU) (n = 99)	1960±3519	
Optimal Cytoreduction		
Optimal (<1cm)	57	51.8
Not optimal	53	48.2
Grade		
Grade 1	26	23.6
Grade 2	41	37.3
Grade 3	43	39.1
Median survival time, months (range)	31 (1–81)	

doi:10.1371/journal.pone.0009615.t001

Identification of PFS-Related Profile

Using Agilent Whole Human Genome Oligo microarray, we generated gene expression data for 110 advanced-stage serous ovarian cancer patients. Then this dataset was used as a discovery set for the identification of PFS-related profile in patients with advanced-stage serous ovarian cancer. To further evaluate the PFS-related profile, we prepared a part of the GSE9891 dataset as an external independent dataset using Affymetrix Human Genome U133 Plus 2.0 Array (See Materials and Methods) [20]. To deal with cross-platform microarray data appropriately, we analyzed only common genes (28304 probes in Agilent platform; 38497 probes in Affymetrix platform) between the two platforms in this study. Of 28304 Agilent probes, 18178 probes with expression levels marked as "Present" in all of the 110 microarray data from the discovery set was further extracted to remove missing and uncertain signals on gene expression, and then the data were per-gene normalized in each dataset by transforming the expression of each gene to a mean of 0 and standard deviation of 1 (Figure S2).

A univariate Cox proportional hazard model showed that expression levels of 97 probes (representing 88 nonredundant genes) were correlated with PFS time ($p<0.01$). In case of multiple-tagged 8 genes (represented by 17 probes), we selected 8 probes (one probe per gene) with the largest sum of the squares of individual expression values for the respective genes as representatives [24]. A total of 88 genes (represented by 88 unique probes) were thereby identified as PFS-related profile. Furthermore, we applied the ridge regression model to estimate optimal regression coefficients (β) for 88 genes of the PFS-related profile (Table 2), and calculated the prognostic index for each sample using equation (1) as reported previously [18]. The 88-gene prognostic indices obtained were in the range of -5.09 to 4.14 (median, 0.11), and the frequency distribution of the indices among 110 patients was unimodal.

To assess the prognostic index as a categorical variable, we attempted to divide this dataset into two groups based on median prognostic index of 0.11 [9]. Patients were assigned to the "high-risk" group if their prognostic index was greater than or equal to the median value, whereas "low-risk" group was composed of cases with the prognostic indices that were less than the median. As shown in Figure 1A, patients with high-risk prognostic indices had shorter median PFS times than those belonging to low-risk group (median PFS, 12 months vs. 51 months; log rank test, $p<0.0001$).

We then performed univariate and multivariate Cox proportional hazard analyses to prove that the 88-gene prognostic index was an independent prognostic factor (Table 3). A univariate Cox's proportional hazard analysis showed that the prognostic index, stage, optimal surgery, and histological grade were correlated with PFS ($p<0.0001$, $p=0.022$, $p<0.0001$ and $p=0.016$, respectively). Moreover, a multivariate analysis showed that the prognostic index was most significantly associated with PFS time [hazard ratio (HR), 3.80; 95% confidence interval (CI), 2.68–5.61; $p<0.0001$].

Validation by Quantitative Real-Time RT-PCR

To validate the microarray expression data, we performed quantitative real-time RT-PCR for a subset of the discovery dataset (53 samples). The four genes, *E2F2*, *FOXJ1*, *DNAH7*, and *FILIP1*, were randomly selected for this purpose. There were significant correlations between microarray expression data and real-time RT-PCR expression data (Figure 2). In spite of the smaller sample size, we confirmed a significant association between PFS time and each of the real-time RT-PCR data for the four genes in the univariate Cox hazard model (data not shown).

Appling PFS-Related Profile to the External Dataset

We translated the 88 prognostic genes with Agilent Probe IDs to Affymetrix 196 probes using a translation function in GeneSpring GX 10 and evaluated the present PFS-related profile in the external dataset (Figure S2). We calculated the prognostic index for each sample in the external dataset by the weighted sum of the expression values of 88 PFS-related genes according to the equation (1), in which the ridge regression coefficients (β) identified in the discovery set were used as weights for the respective genes (See Materials and Methods). We obtained prognostic indices ranging from -5.37 to 4.56 in the external dataset. The frequency distribution of the prognostic indices was not statistically different from that in the discovery set by Kolmogorov Smirnov test ($p>0.05$).

When we divided the external dataset into two subgroups by the median prognostic index (0.11) in the discovery set, a significant correlation was observed between risk classification and PFS (log rank test, $p=0.0004$) (Figure 1B). In univariate analysis of the external data, the estimated prognosis index and optimal surgery were correlated with PFS time ($p=0.0001$ and 0.049, respectively) (Table 3). Multivariate analysis showed that prognostic index was an independent prognostic factor for PFS time (HR, 1.64; 95% CI, 1.27–2.13, $p=0.0001$).

Assessment of Our Prognostic Index

To assess the sensitivity and specificity of our prognostic index, we used ROC curves for the index. An area under ROC curve of 0.5 (indicated by diagonal dotted lines in Figure S3) represents equality between true positive and false positive test results. The extent to which the ROC curve departs from the diagonal line to left and top axes is a measure of the effectiveness of the 88-gene prognostic index in the prediction of clinical outcome. The area under the ROC curves to distinguish early-relapse patients with less than 18 months of PFS times from late-relapse patients was 0.959 and 0.674 in the discovery set and the external dataset, respectively (Figure S3). When we used median value of prognostic index in the discovery set as the cut-off, the sensitivity and specificity were 88.9% and 85.7% in discovery dataset and 64.4% and 69.2% in the external dataset.

We performed survival analysis after the stratification of patients according to the status of debulking surgery which was an independent prognostic factor in multivariate analysis of the discovery dataset (Table 3). We divided patients into two groups ("optimal group" and "suboptimal group") in each of the discovery and external datasets, and assigned each patient to "high-risk" or "low-risk" based on the median value of the current prognostic index in each stratum according to the debulking status. Kaplan-Meier survival analysis showed that high-risk patients had significant shorter PFS time than low-risk patients in each of the four strata from the two datasets (Figure 3) as follows: optimal group ($p<0.0001$) and suboptimal group ($p<0.0001$) in our dataset; optimal group ($p=0.0034$) and suboptimal group ($p=0.015$) in the external dataset. This stratified analysis also indicated that the prognostic index was associated with PFS time independently of the debulking status.

Correlation between This Prognostic Index and Overall Survival

Overall survival is another important endpoint in patients with advance-stage ovarian cancers, and hence we further examined if the present 88-gene prognostic index could be extended to use for predicting the overall survival of patients. To evaluate correlation between this prognostic index and overall survival time, we performed Kaplan-Meier survival curve analysis. Patients with high-risk prognostic indices had shorter overall survival times than the low-risk patients in the two datasets (log rank test, $p<0.0001$ and $p=0.0010$, respectively) (Figure 1C, D). Furthermore, the prognostic index was significantly associated with overall survival time in both the discovery set and the external dataset in multivariate analysis (Table 4).

In addition, we examined the predictive ability of our prognostic index in publicly available Dressman's dataset [25], in which patients were longer followed-up (median overall survival, 31 months; range, 1–185 months). Dressman's dataset [25] was composed of 119 advanced-stage serous ovarian cancer patients treated with platinum-based chemotherapy (including non-taxane chemotherapy). Because their data were generated by a different platform with the foregoing two datasets, 75% of 88 PFS-related genes were translated for survival prediction in this dataset. When we divided Dressman's dataset [25] into two subgroups by the median prognostic index in discovery dataset, a significant association was observed between risk classification and overall survival (log rank test, $p=0.0008$) (Figure S4). Its prognostic index

Table 2. Eighty-eight genes composing the progression-free survival-related profile.

GenBank Acc.	GeneSymbol	Cytoband	β_{ridge}[a]	Description
NM_001123	*ADK*	10q22.2	0.006	adenosine kinase
NM_006408	*AGR2*	7p21.1	0.128	anterior gradient homolog 2 (Xenopus laevis)
NM_080429	*AQP10*	1q21.3	−0.162	aquaporin 10
NM_001040118	*ARAP1*	11q13.4	0.141	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1
NM_006420	*ARFGEF2*	20q13.13	0.032	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)
NM_181575	*AUP1*	2p13.1	0.129	ancient ubiquitous protein 1
NM_004776	*B4GALT5*	20q13.13	0.215	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5
NM_138639	*BCL2L12*	19q13.33	−0.189	BCL2-like 12 (proline rich)
NM_020643	*C11orf16*	11p15.4	0.221	chromosome 11 open reading frame 16
NM_145061	*C13orf3*	13q12.11	−0.107	chromosome 13 open reading frame 3
NM_024032	*C17orf53*	17q21.31	−0.184	chromosome 17 open reading frame 53
NM_001144956	*C1orf230*	1q21.3	0.012	chromosome 1 open reading frame 230
NM_022106	*C20orf177*	20q13.33	0.167	chromosome 20 open reading frame 177
NM_000715	*C4BPA*	1q32.2	−0.505	complement component 4 binding protein, alpha
NM_012337	*CCDC19*	1q23.2	−0.162	coiled-coil domain containing 19
NM_015603	*CCDC9*	19q13.32	0.263	coiled-coil domain containing 9
NM_005408	*CCL13*	17q12	−0.228	chemokine (C-C motif) ligand 13
NM_001252	*CD70*	19p13.3	−0.204	CD70 molecule
NM_078481	*CD97*	19p13.12	−0.137	CD97 molecule
NM_006383	*CIB2*	15q25.1	0.359	calcium and integrin binding family member 2
NM_182848	*CLDN10*	13q32.1	−0.292	claudin 10
NM_001316	*CSE1L*	20q13.13	−0.220	CSE1 chromosome segregation 1-like (yeast)
NM_024295	*DERL1*	8q24.13	0.007	Der1-like domain family, member 1
NM_001042517	*DIAPH3*	13q21.2	0.022	diaphanous homolog 3 (Drosophila)
NM_021120	*DLG3*	Xq13.1	−0.039	discs, large homolog 3 (Drosophila)
NM_020877	*DNAH2*	17p13.1	−0.378	dynein, axonemal, heavy chain 2
NM_018897	*DNAH7*	2q32.3	0.226	dynein, axonemal, heavy chain 7
NM_001394	*DUSP4*	8p21.1	0.007	dual specificity phosphatase 4
NM_004091	*E2F2*	1p36.12	0.220	E2F transcription factor 2
NM_006795	*EHD1*	11q13.1	0.248	EH-domain containing 1
NM_020819	*FAM135A*	6q13	0.142	family with sequence similarity 135, member A
NM_032181	*FAM176A*	2p12	−0.096	family with sequence similarity 176, member A
NM_015687	*FILIP1*	6q14.1	−0.188	filamin A interacting protein 1
NM_021784	*FOXA2*	20p11.21	0.184	forkhead box A2
NM_001454	*FOXJ1*	17q25.1	−0.344	forkhead box J1
NM_000819	*GART*	21q22.11	0.140	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase
NM_178172	*GPIHBP1*	8q24.3	0.147	glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1
NM_000189	*HK2*	2p13.1	−0.087	hexokinase 2
NM_002118	*HLA-DMB*	6p21.32	−0.288	major histocompatibility complex, class II, DM beta
NM_022465	*IKZF4*	12q13.2	−0.092	IKAROS family zinc finger 4 (Eos)
NM_016584	*IL23A*	12q13.2	0.493	interleukin 23, alpha subunit p19
NM_006801	*KDELR1*	19q13.32	−0.001	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1
NM_014895	*KIAA1009*	6q14.3	−0.150	KIAA1009
NM_017527	*LY6K*	8q24.3	0.226	lymphocyte antigen 6 complex, locus K
NM_005906	*MAK*	6p24.2	0.271	male germ cell-associated kinase
NM_024871	*MAP6D1*	3q27.1	−0.038	MAP6 domain containing 1
NM_031417	*MARK4*	19q13.32	0.040	MAP/microtubule affinity-regulating kinase 4
NM_024298	*MBOAT7*	19q13.42	−0.058	membrane bound O-acyltransferase domain containing 7
NM_002421	*MMP1*	11q22.2	−0.336	matrix metallopeptidase 1 (interstitial collagenase)
NM_181526	*MYL9*	20q11.23	0.058	myosin, light chain 9, regulatory

Table 2. Cont.

GenBank Acc.	GeneSymbol	Cytoband	β_{ridge}[a]	Description
NM_032344	*NUDT22*	11q13.1	0.198	nudix (nucleoside diphosphate linked moiety X)-type motif 22
NM_007224	*NXPH4*	12q13.3	−0.310	neurexophilin 4
NM_015311	*OBSL1*	2q35	−0.045	obscurin-like 1
NM_014982	*PCNX*	14q24.2	−0.098	pecanex homolog (Drosophila)
NM_014317	*PDSS1*	10p12.1	0.001	prenyl (decaprenyl) diphosphate synthase, subunit 1
NM_024420	*PLA2G4A*	1q31.1	0.107	phospholipase A2, group IVA (cytosolic, calcium-dependent)
NM_016341	*PLCE1*	10q23.33	0.029	phospholipase C, epsilon 1
NM_001031745	*RIBC1*	Xp11.22	0.209	RIB43A domain with coiled-coils 1
NM_015653	*RIBC2*	22q13.31	0.053	RIB43A domain with coiled-coils 2
NM_006987	*RPH3AL*	17p13.3	−0.043	rabphilin 3A-like (without C2 domains)
NM_001025070	*RPS14*	5q33.1	0.013	ribosomal protein S14
NM_152732	*RSPH9*	6p21.1	−0.102	radial spoke head 9 homolog (Chlamydomonas)
NM_014433	*RTDR1*	22q11.22	0.034	rhabdoid tumor deletion region gene 1
NM_005500	*SAE1*	19q13.32	0.038	SUMO1 activating enzyme subunit 1
NM_020150	*SAR1A*	10q22.1	0.277	SAR1 homolog A (S. cerevisiae)
NM_031469	*SH3BGRL2*	6q14.1	−0.281	SH3 domain binding glutamic acid-rich protein like 2
NM_003951	*SLC25A14*	Xq25	−0.344	solute carrier family 25 (mitochondrial carrier, brain), member 14
NM_014585	*SLC40A1*	2q32.2	0.065	solute carrier family 40 (iron-regulated transporter), member 1
NM_052910	*SLITRK1*	13q31.1	−0.314	SLIT and NTRK-like family, member 1
NM_172312	*SPAG8*	9p13.3	−0.123	sperm associated antigen 8
NM_145263	*SPATA18*	4q12	0.041	spermatogenesis associated 18 homolog (rat)
NM_006100	*ST3GAL6*	3q12.1	−0.192	ST3 beta-galactoside alpha-2,3-sialyltransferase 6
NM_018414	*ST6GALNAC1*	17q25.1	−0.175	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3) -N-acetylgalactosaminide alpha-2,6-sialyltransferase 1
NM_032872	*SYTL1*	1p36.11	−0.084	synaptotagmin-like 1
NM_014466	*TEKT2*	1p34.3	−0.226	tektin 2 (testicular)
NM_005424	*TIE1*	1p34.2	0.250	tyrosine kinase with immunoglobulin-like and EGF-like domains 1
NM_198276	*TMEM17*	2p15	0.025	transmembrane protein 17
NM_199203	*TMEM189 -UBE2V1*	20q13.13	0.174	TMEM189-UBE2V1 readthrough transcript
NM_033550	*TP53RK*	20q13.12	0.054	TP53 regulating kinase
NM_139075	*TPCN2*	11q13.2	0.034	two pore segment channel 2
NM_018430	*TSNAXIP1*	16q22.1	0.170	translin-associated factor X interacting protein 1
NM_014023	*WDR37*	10p15.3	0.296	WD repeat domain 37
NM_018053	*XKR8*	1p35.3	0.106	XK, Kell blood group complex subunit-related family, member 8
NM_015896	*ZMYND10*	3p21.31	0.052	zinc finger, MYND-type containing 10
NM_005773	*ZNF256*	19q13.43	0.048	zinc finger protein 256
NM_024691	*ZNF419*	19q13.43	−0.042	zinc finger protein 419
NM_021089	*ZNF8*	19q13.43	0.093	zinc finger protein 8
NM_017975	*ZWILCH*	15q22.31	−0.074	Zwilch, kinetochore associated, homolog (Drosophila)

[a]A regression coefficient of each gene in ridge regression extension of multivariate Cox hazard model.

doi:10.1371/journal.pone.0009615.t002

was significantly correlated with overall survival time in multivariate analysis (HR, 1.51; 95% CI, 1.19–1.93, $p = 0.0008$).

Characterization of PFS-Related Profile

We conducted GO analysis to understand the biological characteristics of 88 PFS-related genes. To characterize the gene list based on GO classification on 'biological process', 'molecular function', and 'cellular component', we examined which categories were highly associated with the 88 genes. After multiple testing corrections using the FDR method [26], 8 categories were significantly overrepresented (FDR q-value<0.10) (Figure 4). In the 88 PFS-related genes, genes involved in GTPase binding (GO17016, GO31267 and GO51020), cellular localization (GO51649 and GO51641), intracellular transport (GO46907 and GO6886), and/or ciliary or flagellar motility (GO1539) were notably enriched. We investigated similarities in overrepresented GO categories between our 88 PFS-related genes and the previously reported gene expression profiles which were correlated to prognosis in ovarian cancer [11,13]. However, we could not identify common GO categories between our profile and the previously reported profiles (data not shown).

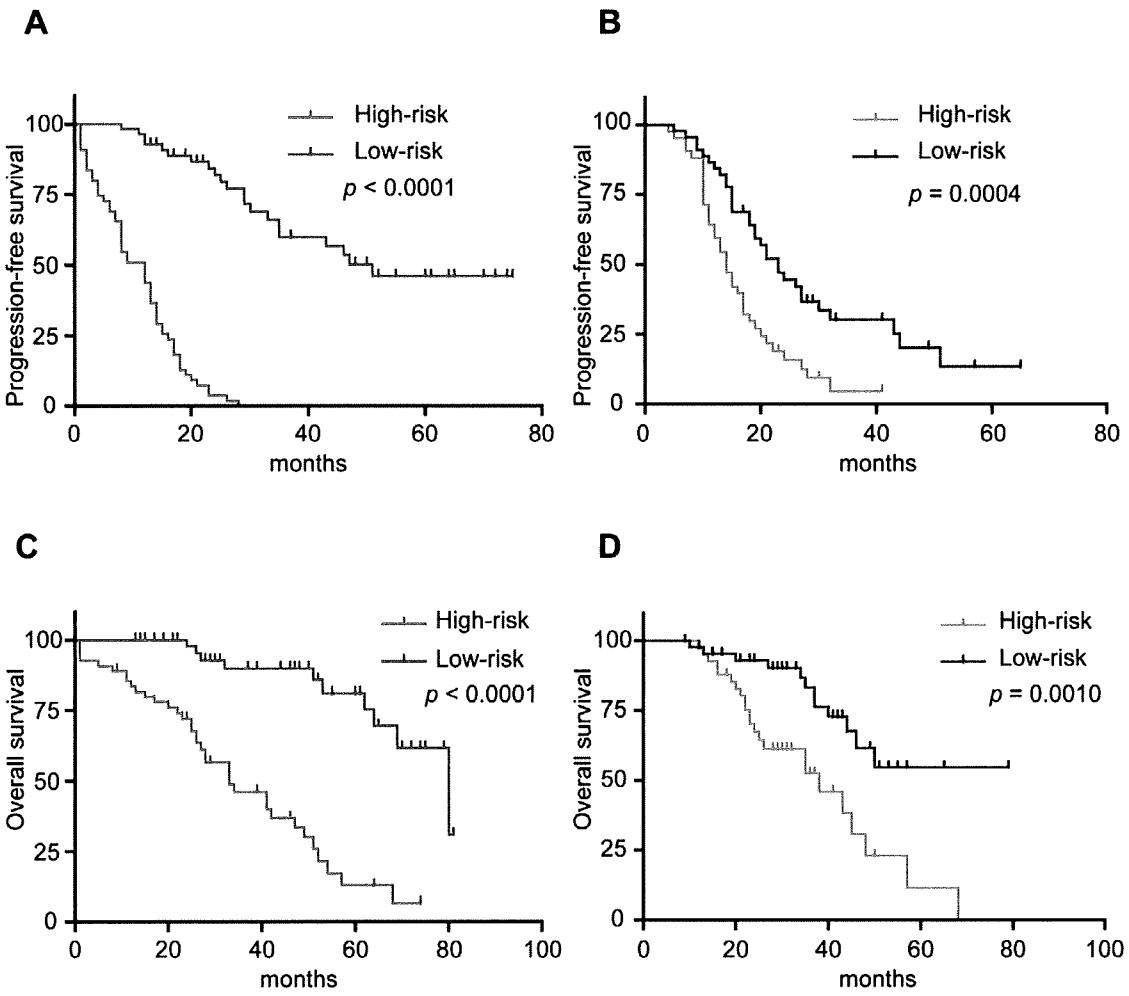

Figure 1. Prediction of prognosis in high-risk and low-risk patients based on the prognostic index. High-risk patients had significantly short progression-free survival times compared to low-risk patients (A) in the discovery set (log rank test, $p<0.0001$) and (B) in the external set (log rank test, $p=0.0004$). Similarly, high-risk patients had significantly shorter overall survival times compared to low-risk patients (C) in the discovery set (log rank test, $p<0.0001$) and (D) in the external set (log rank test, $p=0.0010$).
doi:10.1371/journal.pone.0009615.g001

We further used IPA software to analyze 88 PFS-related genes from the viewpoint of molecular interaction or pathway. Top three significant networks (score>25) are shown in Figures S5-7. The network 1 included 15 of the 88 prognostic genes, and was significantly associated with IPA-defined several networks: cell death, neurological disease, and cellular assembly and organization (Figure S5). Fourteen prognostic genes were included in the network 2, which was defined as networks related to cancer, cell morphology, and renal and urological disease (Figure S6). The network 3 displayed significant interactions and interrelations between genes involved in cell-to-cell signaling and interaction, hematological system development and function, and immune cell trafficking (Figure S7). In the 88 genes, we found several genes interacting with *SRC* or *MYC* (Figure S6), each of which was reported as a representative gene in oncogenic pathways of ovarian cancer [25,27].

Discussion

In this study, we identified the prognostic index for predicting PFS time in patients with advanced-stage serous ovarian cancer treated with platinum/taxane-based adjuvant chemotherapy across two types of microarray expression data from the present discovery set and publicly available external set by using the ridge regression Cox model. The significant correlation between our prognostic index and OS time was also indicated in the two independent datasets.

In expression microarray analysis, there is a so-called "curse of dimensionality" problem that the number of genes is much larger than the number of samples. To improve the reliability of a gene expression-based prognostic model, it is necessary to avoid overfitting to the dataset, and to confirm the reproducibility of the predictive ability in external independent datasets [28]. Until now, several bioinformatics approaches have been proposed to establish a model for survival prediction using microarray data [18,29]. Bøvelstad *et al.* [18] recently examined the prediction performance of the following seven methods: univariate selection, forward stepwise selection, principal components regression, supervised principal components regression, partial least squares regression, ridge regression and the lasso using three microarray datasets [Dutch breast cancer data (n = 295), diffuse large B-cell lymphoma data (n = 240), and Norway/Stanford breast cancer data (n = 115)] [7,30–32]. They concluded that the univariate Cox model alone was insufficient for predicting survival and that the ridge regression