

ORIGINAL ARTICLE

Microsatellite typing for DRB1 alleles: application to the analysis of HLA associations with rheumatoid arthritis

H-S Lee^{1,6}, W Li¹, A Lee¹, P Rodine², RR Graham^{3,4}, WA Ortmann², F Batliwalla¹, KW Lee⁵, SC Bae⁶, TW Behrens² and PK Gregersen¹

¹Feinstein Institute for Medical Research, North Shore LIJ Health System, Manhasset, NY, USA; ²Department of Medicine, Division of Rheumatic and Autoimmune Diseases, University of Minnesota Medical School, Minneapolis, MN, USA; ³Program in Medical and Population Genetics, Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, MA, USA; ⁴The Center for Human Genetics Research and Molecular Biology, Massachusetts General Hospital, Boston, MA, USA; ⁵Hallym Institution for Genome Application, Hallym University Sacred Heart Hospital, Anyang, South Korea and ⁶Hanyang University College of Medicine, Hospital for Rheumatic Diseases, Seoul, South Korea

*The current methods for molecular typing of HLA-DR alleles incur a substantial financial burden when performing large population studies. In the current study, we aimed to provide much less expensive typing approach with high predictability for DRB1 genotype. We have used a panel of three microsatellite markers in the class II region (D6S2666, D6S2665 and D6S2446) for genotyping and haplotype reconstruction in a total of 1687 Caucasian (1313 RA patients and 374 controls) and 1364 Korean individuals (744 RA patients and 620 controls), all of whom were previously genotyped for DRB1. We found that a total of 88.4 and 87.4% of all observed three-marker haplotypes could determine the DR type with a positive predictive value > 0.8 with high sensitivity and specificity. There was a high degree of haplotype conservation when comparing Caucasian and Asian populations. Interestingly, we found that the majority of DRB1*09 and DRB1*10 alleles share a common three-marker haplotype in both Caucasian and Asian populations. This is unexpected, since these two alleles are found on very different haplotype families. In addition, these two alleles are both associated with rheumatoid arthritis, making the elucidation of these haplotype relationships potentially important for understanding disease susceptibility.*

Genes and Immunity (2006) 7, 533–543. doi:10.1038/sj.gene.6364325; published online 20 July 2006

Keywords: microsatellite typing; HLA-DRB1; rheumatoid arthritis

Introduction

The major histocompatibility complex (MHC) class II region encodes α - and β -chains of the DR, DQ and DP isotypes. These heterodimeric molecules are expressed on the cell surface and have a primary role in presenting peptide antigens to T cells. These loci are characterized by extensive polymorphism that is concentrated on the amino-acid residues that shape the peptide binding site and this allelic diversity has been associated with a wide variety of autoimmune, inflammatory and infectious diseases.^{1,2} The exact mechanism underlying these associations is for the most part unknown.^{3,4} Furthermore, the patterns of association are quite complex and despite the large literature, gaps in our knowledge remain. It is apparent that very large population studies continue to be a valuable means of understanding the diverse ways in which the MHC contributes to disease risk.

The current standard for typing class II alleles involve the use of sequence specific oligonucleotide probes (SSOP), also known as 'oligotyping'.⁵ In the case of DR loci, this method involves PCR amplification of specific exonic regions (primarily exons 2 and 3) of DR β using multiple primer pairs and hybridizing the amplified material with sequence-specific probes. While quite precise, the method is labor intensive and expensive, with costs ranging up to hundreds of dollars per sample, depending on the resolution required. This makes large population studies a significant financial burden.

In the current report, we have used a panel of three microsatellites in the class II regions to provide a surrogate for traditional DRB1 oligotyping. This typing approach may be particularly useful for a 'first pass' at characterizing large population samples, at a cost of less than \$1 per sample. Selected samples may then be subjected to more refined class II typing. In the course of these studies, we have demonstrated remarkable conservation of some microsatellite-DRB1 haplotypes, and many of these are conserved across Caucasian and Asian populations. Finally, an unexpectedly close relationship between DR9 and DR10 haplotypes has been observed. This finding may have relevance to understanding the MHC associations with RA, which have recently been shown to involve several distinct loci within the MHC.

Correspondence: Professor PK Gregersen, Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, NY 11030, USA.
E-mail: peterg@nshs.edu
Received 6 April 2006; revised 12 June 2006; accepted 12 June 2006; published online 20 July 2006

Results

HLA-DRB1 typing and reconstruction of three-marker microsatellite haplotypes in Caucasian and Korean populations

A total of 1687 Caucasian and 1364 Korean DNA samples were used for the analysis. These samples included 1313 Caucasian patients with RA and 374 Caucasian controls; the Korean samples consisted of 744 RA patients and 620 controls. Supplementary Table 1 shows the DRB1 allele distribution in these populations, listed in order of frequency in controls. A total of 11 DR types and nine DRB1*04 subtypes were defined in Caucasians. In the Korean population, full DR typing was carried out for all allelic groups, leading to the definition of 36 DR subtypes. The frequencies of DRB1 in these populations are similar to previous reports, with enrichment of 'shared epitope' alleles in both RA patient groups.

In addition to DRB1 genotypes, three microsatellite markers in the class II region were analyzed, as shown in Figure 1. These markers had primer pairs flanking bp positions (Build 35) 32505448 through 32505591 (D6S2666), 32511335 through 32511577 (D6S2665) and 32721575 through 32721726 (D6S2446). These microsatellite markers were found previously to be very polymorphic,^{6,7} and that was confirmed here: D6S2666, nine alleles; D6S2665, 10 alleles; and D6S2446, 16 alleles. A number (e.g. 6-7-3) was assigned for each allele at each marker, and haplotypes based only on these three markers were reconstructed using the PHASE program.^{8,9} Table 1 shows the most frequent three-marker

haplotypes (>1%) observed in the control populations of Caucasians and Koreans. Approximately 86% (Caucasian) and 84% (Korean) of all haplotypes in the population are represented in this group of most frequent haplotypes. The frequency of these same three-marker haplotypes in the RA patients is also shown for comparison in Table 1. Not surprisingly, several haplotypes show significant frequency differences between cases and controls in each ethnic group. In Caucasians, the '6-7-3' and '6-8-3' haplotypes show the greatest enrichment in RA vs Controls with odds ratios (OR) of 3.68 (95% CI 2.86–4.72) and 6.41 (95% CI 3.38–12.16) respectively. In contrast, the '1-1-3' and '6-5-5' haplotypes are most associated with RA in the Korean



Figure 1 The location of three microsatellite markers (D6S2666; 32505448-32505591, D6S2665; 32511335-32511577 and D6S2446; 32721575-32721726) relative to the HLA-DRB1 (32654527-32665559).

Table 1 The distribution of three microsatellite marker haplotypes in Caucasian and Korean populations

Haplotype ^a	Caucasians				Koreans				
	Controls		RA patients		Controls		RA patients		
	Number	% ^b	Number	% ^b	Number	% ^b	Number	% ^b	
116	97	13.0	198	7.5	673	118	9.5	126	8.5
673	77	10.3	779	29.7	113	116	9.4	317	21.3
652	76	10.2	112	4.3	655	112	9.0	244	16.4
368	61	8.2	208	7.9	116	105	8.5	93	6.3
487	60	8.0	310	11.8	487	88	7.1	120	8.1
1116	43	5.7	59	2.2	559	70	5.6	54	3.6
111	29	3.9	55	2.1	2410	66	5.3	11	0.7
1114	24	3.2	50	1.9	746	63	5.1	37	2.5
1110	21	2.8	28	1.1	119	52	4.2	39	2.6
2411	19	2.5	30	1.1	652	43	3.5	31	2.1
115	18	2.4	40	1.5	1110	26	2.1	31	2.1
666	15	2.0	9	0.3	6715	26	2.1	12	0.8
655	14	1.9	87	3.3	6716	24	1.9	23	1.5
117	14	1.9	25	1.0	646	23	1.9	11	0.7
746	14	1.9	15	0.6	653	21	1.7	50	3.4
6716	13	1.7	16	0.6	345	21	1.7	18	1.2
683	10	1.3	210	8.0	115	19	1.5	18	1.2
647	10	1.3	32	1.2	245	18	1.5	13	0.9
645	10	1.3	17	0.6	745	18	1.5	9	0.6
3511	9	1.2	10	0.4	2411	13	1.0	30	2.0
6511	9	1.2	9	0.3					
Total	643	86.0	2299	87.5	Total	1042	84.0	1287	86.5

The most frequent haplotypes (>1%) in controls are listed in each population.

^aEach digit means an allele of each three markers; D6S2666, D6S2665 and D6S2446 in order. In cases of four digits, the last two digits represent an allele of D6S2446.

^b% means the percentage of the haplotypes among all haplotypes in each group.

population with OR = 2.62 (95% CI 2.09–3.29) and OR = 1.97 (95% CI 1.56–2.51) respectively.

Reconstruction of four marker haplotypes at D6S2666-D6S2665-DRB1-D6S2446

We then used the PHASE program^{8,9} to reconstruct four-marker haplotypes, including the DRB1 genotypic data along with the three microsatellite markers D6S2666, D6S2665 and D6S2446. A total of 173 and 149 four-marker haplotypes were reconstructed in the Caucasian and Korean populations, respectively (Table 2 and 3). Although a large number of haplotypes could be identified, many of these were rare (frequency <0.1%). In the Caucasian population, a little over one third of the total haplotypes 62/173 (35.8%) had a frequency >0.1%, and these accounted for 95.2% of the haplotypes found in the Caucasian population. In the Korean population, a similar pattern was observed in that 74/149 (50%) of the haplotypes had frequencies of >0.1% and these accounted for the majority (96.7%) of the haplotypes in the Korean population.

The four-marker haplotypes shown in Tables 2 and 3 are grouped by DRB1 alleles or allelic families. Note that for each DRB1 group, only a few haplotypes tend to dominate. For example, the '4-8-DRB1*01-7' combination accounts for 85% of the DR1-associated haplotypes in Caucasians. DRB1*15/16, *03, *10, *11 and *12 show similar enrichment of a predominant haplotype within the group. Other DRB1 allelic families, such as DRB1*07 and DRB1*13, have several common haplotypes within the group. A similar pattern is observed in the haplotypes associated with the various Korean DRB1 alleles.

Note that within the DRB1*04 group of alleles, the '6-7-3' and '6-8-3' microsatellite haplotypes predominate, being enriched particularly in the DRB1*0401 and 0404 allelic subsets. This is consistent with the association of these three-marker haplotypes with RA, as shown in Table 1. Similarly, the most strongly RA associated three-marker haplotypes in Koreans, '1-1-3' and '6-5-5' are predominantly found in association with DRB1*0405 and DRB1*0901, respectively. This is consistent with HLA associations with RA that have been previously reported in the Korean population.

Three microsatellite markers can be used to predict the HLA-DRB1 type

To determine whether particular haplotypes of three markers can be useful to predict DRB1 type or subtype, we obtained the sensitivity, specificity, positive predictive value, negative predictive value and likelihood ratio for all haplotypes with greater than a 0.1% frequency. Tables 4 and 5 showed the data of those parameters for DRB1 types other than DRB1*04. Tables 6 and 7 shows these data for the DRB1*04 subtypes in each population. In addition to the most common haplotypes, we have included all the haplotypes with positive predictive values more than 0.8 in order to show the haplotypes that can be useful for DR typing in each population.

Despite high positive predictive values, many haplotypes within a DR group displayed low-sensitivity, because multiple haplotypes may be associated with a particular DRB1 allele. Thus, for example, as shown in Table 4, DRB1*03 contains one major ('3-6-8') and two minor haplotypes in Caucasians. The minor microsatel-

lite haplotypes ('6-4-5' and '6-4-6') have low-sensitivity but high-positive predictive value. However, if any of these three haplotypes are present, the sensitivity is high (0.927) and the PPV is close to 1 (0.997). Thus, in a practical sense, combinations of three-marker haplotypes can be used as good predictors of many of the common HLA-DRB1 allelic groups.

In one of our previous studies using the same three microsatellite markers, we reported that the particular haplotype '1-1-6' was nearly 100% accurate in predicting the haplotype 'DRB1*1501-DQB1*0602'.⁷ Even though this three-marker haplotype was the major one predicting DRB1*15 in our current study, there were the other minor ones in both the Korean and Caucasian populations. This slight difference in accuracy between the two studies undoubtedly relates to the fact that several DRB1*1501 haplotypes exist in the population¹⁰ other than the 'DRB1*1501-DQB1*0602' haplotype that is associated with systemic lupus and that was the focus of our previous analysis.⁷

Within the DR4 group of alleles, typing specificity is less reliable as an indicator of DR4 subtype, particularly in Caucasians (Table 6). This is because many three-marker microsatellite haplotypes are in LD with several of the DR4 subtypes. For example, the '6-7-3' haplotype is commonly found in association with DRB1*0401, 0403 and 0404 alleles. At the same time, while the '6-8-3' haplotype is quite specific (0.995) and predictive (0.945) for DRB1*0401, it is not very sensitive (0.28) as it is only found on a minority of 0401 haplotypes. The situation in Korean population is somewhat less complex, in part, because the DRB1*0405 allele is predominant in this population, as shown in Table 7. Thus, as shown in Table 7, the '1-1-3' haplotypes is highly specific (0.99) and quite sensitive (0.88) for the *0405 allele, with a PPV = 0.95.

A striking feature of these data is that the DRB1*09 and DRB1*10 alleles were not well delineated using this three-marker haplotype method. This is because a large majority of DRB1*09 and almost all DRB1*10 alleles contained the haplotype '6-5-5' three-marker microsatellite haplotype. This was true in both Korean and Caucasian populations. As discussed below, this haplotype similarity is unexpected, since DR9 and DR10 belong to quite distinct haplotype families.^{10–14}

Discussion

We have established a microsatellite haplotyping method that can serve as a surrogate for 'broad level' DRB1 genotyping with very good sensitivity and specificity for most of the major DRB1 allelic groups. By using three microsatellite markers, we found that a total of 88.4 and 87.4% of all observed three-marker haplotypes could determine the DR type with a positive predictive value >0.8 in Caucasian and Korean populations, respectively. This method will be especially useful for DR screening of large populations, since it can be carried out at a fraction of the cost of traditional DRB1 oligotyping. For studies that require more definitive DRB1 subtyping, investigators can select samples that carry the haplotypes common to particular allelic groups, and focus the more expensive oligotyping on those samples that require precise genetic definition.

Table 2 The distribution of all haplotype combinations according to HLA-DRB1 in Caucasian populations (The total number of subjects are 1687 and the total number of haplotypes are 3374)

DR	Haplotype ^a	Number	% ^b	DRB1	Haplotype ^a	Number	% ^b	DRB1	Haplotype ^a	Number	% ^b				
01	487	370	10.97	*0404/*0408 ^c	873	2	0.06	11	655	1	0.03				
	117	32	0.95		111	1	0.03		662	1	0.03				
	482	8	0.24		112	1	0.03		662	1	0.03				
	116	6	0.18		623	1	0.03		672	1	0.03				
	497	6	0.18		693	1	0.03		772	1	0.03				
	477	2	0.06	*0405	663	50	1.48	12	1511	1	0.03				
	4107	2	0.06		673	6	0.18		111	55	1.63				
	187	1	0.03		113	4	0.12		115	1	0.03				
	366	1	0.03	07	563	1	0.03	13	246	1	0.03				
	367	1	0.03		1116	101	2.99		655	1	0.03				
	387	1	0.03		1114	74	2.19		673	1	0.03				
	485	1	0.03		6716	29	0.86		2411	47	1.39				
	673	1	0.03		1113	15	0.44		3512	27	0.80				
	118	1	0.03		6616	13	0.39		111	23	0.68				
	03	368	268	7.94	08	113	7	0.21	14	1110	20	0.59			
		645	27	0.80		112	6	0.18		3511	19	0.56			
		646	11	0.33		673	5	0.15		6711	16	0.47			
		358	3	0.09		1115	5	0.15		6511	13	0.39			
366		3	0.09	6516		4	0.12	652		3	0.09				
117		2	0.06	117		3	0.09	2511		3	0.09				
168		2	0.06	671		3	0.09	641		2	0.06				
348		2	0.06	6717		3	0.09	662		2	0.06				
468		2	0.06	116		2	0.06	2412		2	0.06				
855		2	0.06	657		2	0.06	4715		2	0.06				
955		2	0.06	1112		2	0.06	4812		2	0.06				
118		1	0.03	111		1	0.03	6611		2	0.06				
367		1	0.03	115		1	0.03	118		1	0.03				
658		1	0.03	672		1	0.03	242		1	0.03				
3610		1	0.03	677		1	0.03	1111		1	0.03				
6410		1	0.03	1117		1	0.03	1112		1	0.03				
6416		1	0.03	1314		1	0.03	1312		1	0.03				
*0401		673	491	14.55		09	6615	1		0.03	10	2410	1	0.03	
		683	208	6.16			6715	1		0.03		3412	1	0.03	
		773	14	0.41			11	115		54		1.60	3610	1	0.03
		113	8	0.24				116		5		0.15	3612	1	0.03
	693	6	0.18	555	5			0.15	3712	1		0.03			
	173	4	0.12	114	1			0.03	4712	1		0.03			
	6103	3	0.09	676	1			0.03	4811	1		0.03			
	573	2	0.06	678	1			0.03	746	29		0.86			
	653	2	0.06	10	655		30	0.89	116	12		0.36			
	686	2	0.06		673		6	0.18	358	8		0.24			
	473	1	0.03		563		2	0.06	353	2		0.06			
	663	1	0.03		113		1	0.03	366	2		0.06			
666	1	0.03	118		1	0.03	348	1	0.03						
973	1	0.03	*0402		655	69	2.05	488	1	0.03					
*0402	666	17		0.50	654	4	0.12	676	1	0.03					
	656	1		0.03	755	2	0.06	756	1	0.03					
*0403/*0407 ^c	673	36	1.07	11	652	185	5.48	2 (*15/16)	116	269	7.97				
	666	6	0.18		641	5	0.15		647	42	1.25				
	683	3	0.09		6511	5	0.15		1110	29	0.86				
	653	2	0.06		111	4	0.12		6410	5	0.15				
	663	1	0.03		112	3	0.09		117	2	0.06				
*0404/*0408 ^c	673	310	9.19	11	115	2	0.06	11	747	2	0.06				
	653	32	0.95		742	2	0.06		856	2	0.06				
	663	18	0.53		2411	2	0.06		1116	1	0.03				
	113	15	0.44		116	1	0.03		366	1	0.03				
	683	9	0.27		352	1	0.03		648	1	0.03				
	353	4	0.12		368	1	0.03								

^aEach digit represents an allele of each three markers; D6S2666, D6S2665 and D6S2446, in order. In cases of four digits, the last two digits represent the D6S2446 allele.

^bValues are % of number with the haplotype among a total of 3374 haplotypes.

^cSeveral alleles of DRB1*04 were grouped together when they shared haplotypes.

Shaded area indicates the haplotypes with frequency >0.1%, and these comprise 95.2% of all haplotypes in the population.

As HLA-DR alleles are strong risk factors for many autoimmune, infectious and inflammatory disorders,^{1,2} it is often necessary to take this into account when

studying both the genetic and environmental contributions to disease in populations. This is usually done in the context of either case-control or longitudinal cohort

Table 3 The distribution of all haplotype combinations according to HLA-DRB1 in Korean populations (The total number of subjects are 1364 and the total number of haplotypes are 2728)

DRB1	Haplotype ^a	Number	% ^b	DRB1	Haplotype ^a	Number	% ^b	DRB1	Haplotype ^a	Number	% ^b
*0101	487	208	7.625	*0701	1117	1	0.037	*1202	755	2	0.073
	477	2	0.073		1116	1	0.037		676	1	0.037
	673	1	0.037		1115	1	0.037		675	1	0.037
	657	1	0.037	*0802	115	16	0.587	356	1	0.037	
	485	1	0.037		113	14	0.513	347	1	0.037	
*0301	116	1	0.037	116	11	0.403	113	1	0.037		
	646	34	1.246	126	3	0.11	*1301/1302/1307 ^c	2410	76	2.786	
	642	1	0.037	246	2	0.073		2411	43	1.576	
*0401	673	26	0.953	*0803	559	124	4.545	119	43	1.576	
	683	2	0.073		5510	23	0.843	1110	40	1.466	
	113	2	0.073	555	17	0.623	6715	1	0.037		
	773	1	0.037	116	1	0.037	6710	1	0.037		
*0403/0407 ^c	653	1	0.037	*0901	655	284	10.411	652	1	0.037	
	673	57	2.089		115	14	0.513	*1401	746	51	1.87
	663	4	0.147		645	8	0.293		656	4	0.147
	653	4	0.147		675	6	0.22	747	3	0.11	
*0404/0408 ^c	113	1	0.037	665	2	0.073	745	3	0.11		
	673	26	0.953	755	1	0.037	743	1	0.037		
	653	8	0.293	745	1	0.037	*1402/1403/1404 ^c	458	15	0.55	
	683	7	0.257	653	1	0.037		748	4	0.147	
*0405/0445 ^c	663	1	0.037	555	1	0.037	742	2	0.073		
	655	1	0.037	*1001	655	63	2.309	746	1	0.037	
	113	1	0.037		*1101/1104/1106 ^c	652	72	2.639	459	1	0.037
	113	411	15.066	642		23	0.843	457	1	0.037	
	653	35	1.283	672		6	0.22	*1405	746	47	1.723
	673	10	0.367	242		2	0.073		745	10	0.367
	115	3	0.11	658	1	0.037	748	1	0.037		
	116	2	0.073	645	1	0.037	246	1	0.037		
683	1	0.037	643	1	0.037	*1406/1407/1412 ^c	118	15	0.55		
652	1	0.037	2410	1	0.037		116	5	0.183		
*0406	673	119	4.362	*1201	245	31	1.136	458	3	0.11	
	113	3	0.11		241	22	0.806	746	1	0.037	
	243	2	0.073		351	15	0.55	123	1	0.037	
	683	1	0.037		248	13	0.477	*1501	116	177	6.488
	653	1	0.037		243	8	0.293		356	12	0.44
*0410	653	21	0.77	741	7	0.257	126	7	0.257		
	673	3	0.11	641	3	0.11	111	7	0.257		
*0701	633	2	0.073	111	2	0.073	115	4	0.147		
	6716	47	1.723	658	1	0.037	246	1	0.037		
	6715	37	1.356	648	1	0.037	117	1	0.037		
	6815	21	0.77	361	1	0.037	1110	1	0.037		
	1114	14	0.513	358	1	0.037	*1502	119	48	1.76	
	6714	6	0.22	356	1	0.037		1113	16	0.587	
	6616	5	0.183	352	1	0.037	1110	16	0.587		
	6814	3	0.11	246	1	0.037	669	1	0.037		
	673	2	0.073	238	1	0.037	659	1	0.037		
	671	1	0.037	*1202	345	39	1.43	116	1	0.037	
	6713	1	0.037		745	13	0.477	*1602	647	29	1.063
6710	1	0.037	655		8	0.293	657		2	0.073	
246	1	0.037	365	6	0.22						

^aEach digit represents an allele of each three markers; D6S2666, D6S2665, and D6S2446, in order. In cases of four digits, the last two digits represent the D6S2446 allele.

^bValues are % of number with the haplotype among a total of 2728 haplotypes.

^cSeveral alleles among the same DRs were grouped together when they shared haplotypes.

Shadowed area indicates the haplotypes with frequency >0.1%, and these comprise 96.7% of all haplotypes in the population.

studies. It is our hope that this new approach to HLA-DR typing can make such studies more affordable to a wider array of research laboratories by dramatically reducing the costs of obtaining basic information on DRB1 genotypes. The future analysis and addition of SNPs in the region, in combination with these microsatellite markers, will undoubtedly improve the accuracy of this approach.

Beyond the practical implications, our studies have yielded insights into class II haplotype relationships with potential implications for understanding the MHC associations with rheumatoid arthritis. In general, we found that there is a high degree of haplotype conservation when comparing Caucasian and Asian populations. Some of these shared haplotypes are predominant in both populations, such as the '4-8-DRB1*01-7' haplotype

Table 4 The distribution of highly predictable haplotypes for each HLA-DRB1 type other than DRB1*04 subtypes in Caucasian populations

DRB1 type	Haplotype	Number	SS ^a	SP ^b	PPV ^c	NPV ^d	LR ^e
*01	487	370	0.855	1.000	1.000	0.979	ND
	117	32	0.074	0.998	0.821	0.880	31.1
	482	8	0.018	1.000	1.000	0.874	ND
	497	6	0.014	1.000	1.000	0.873	ND
	All above haplotypes ^f	416	0.961	0.998	0.983	0.994	403.6
*03	368	268	0.812	1.000	0.996	0.980	2472.1
	645	27	0.082	1.000	1.000	0.909	ND
	646	11	0.033	1.000	1.000	0.905	ND
	All above haplotypes ^f	306	0.927	1.000	0.997	0.992	2822.6
*07	1116	101	0.358	1.000	0.990	0.945	1107.4
	1114	74	0.262	1.000	1.000	0.937	ND
	6716	29	0.103	1.000	1.000	0.924	ND
	1113	15	0.053	1.000	1.000	0.921	ND
	6616	13	0.046	1.000	1.000	0.920	ND
	1115	5	0.018	1.000	1.000	0.918	ND
	6516	4	0.014	1.000	1.000	0.918	ND
	All above haplotypes ^f	232	0.823	1.000	0.996	0.984	2543.8
*08	115	54	0.806	0.999	0.931	0.996	666.3
	555	5	0.075	1.000	1.000	0.982	ND
	All above haplotypes ^f	59	0.881	0.999	0.937	0.998	728.0
*09	655	30	0.750	0.979	0.297	0.997	35.2
*10	655	69	0.920	0.990	0.683	0.998	94.8
	654	4	0.053	1.000	1.000	0.979	ND
*09/10	655	99	0.861	0.999	0.980	0.995	1402.8
*11	652	184	0.852	0.999	0.979	0.990	672.5
*13	2411	47	0.241	0.999	0.959	0.955	383.1
	3512	27	0.138	1.000	1.000	0.950	ND
	3511	19	0.097	1.000	1.000	0.948	ND
	6711	16	0.082	1.000	1.000	0.947	ND
	All above haplotypes ^f	109	0.559	0.999	0.982	0.974	888.5
*14	746	29	0.509	1.000	1.000	0.992	ND
*15/16 (DR2)	116	269	0.7599	0.9914	0.9119	0.9724	88.3
	647	32	0.1186	1.0000	1.0000	0.9064	ND
	6410	5	0.0141	0.9997	0.8333	0.8964	42.7
	All above haplotypes ^f	316	0.893	0.991	0.921	0.987	99.8

This table includes the haplotypes with high predictive value >0.8 for DRs among all haplotypes with statistically significantly associated with the particular DRs (All *P*-values less than 0.01 by Fisher's Exact Probability Test), except for [6-5-5] shared by *09 and *1001.

^aSensitivity.

^bSpecificity.

^cPositive predictive value.

^dNegative predictive value.

^eLikelihood ratio.

^fMeans all above haplotypes of given subtype.

ND, not available due to calculation divided by 0.

and the '1-1-DRB1*1501 (DR2)-6' haplotype (see Tables 4 and 5). DR4-related haplotypes also commonly share the three-marker microsatellite combination '6-7-3' in both populations, although the '1-1-DRB1*0405-3' haplotype predominates in the Korean population. This commonality of haplotypes across the major ancestral groups is consistent with the ancient origin of these haplotype families.¹⁵⁻¹⁷ We also observed that certain microsatellites exhibited multiple repeat differences within a haplotype group. For example, multiple alleles of the D6S2446 microsatellite are found on DR*07 haplotypes, involving GT repeats between 13 and 16, giving rise to multiple three-marker haplotypes in this group (e.g. 1113, 1114,

1115, 1116). This may relate to the decreased stability of longer repeats that has been described,¹⁸ and therefore may not reflect a true excess of overall genetic heterogeneity within the DR7 haplotype family.

However, the three-marker haplotype relationship between DR9 and DR10 alleles is unexpected. The majority of DRB1*09 and DRB1*10 alleles share a common three-marker microsatellite combination, '6-5-5' in both Caucasian and Asian populations. This might suggest that DR9 and DR10 belong to a common haplotype family. However, it is quite clear that this is not the case. Overall, the DR subregion has been characterized by missing or alternative arrangements of

Table 5 The distribution of highly predictable haplotypes for each HLA-DRB1 type/subtype other than DRB1*04 subtypes in Korean populations

<i>DRB1 type</i>	<i>DRB1 subtype</i>	<i>Haplotype</i>	<i>Number</i>	<i>SS^a</i>	<i>SP^b</i>	<i>PPV^c</i>	<i>NPV^d</i>	<i>LR^e</i>	
*01	0101	487	208	0.972	1.000	1.000	0.998	ND	
*03	0301	646	34	0.971	1.000	1.000	1.000	ND	
*07	0701	6716	47	0.331	1.000	1.000	0.965	ND	
		6715	37	0.261	1.000	0.974	0.961	673.8	
		6815	21	0.148	1.000	1.000	0.955	ND	
		1114	14	0.099	1.000	1.000	0.953	ND	
		6714	6	0.042	1.000	1.000	0.950	ND	
		6616	5	0.035	1.000	1.000	0.950	ND	
		6814	3	0.021	1.000	1.000	0.949	ND	
		All above haplotypes ^f	133	0.937	1.000	0.993	0.997	2422.1	
		*08	0803	559	124	0.752	1.000	1.000	0.984
5510	23			0.139	1.000	1.000	0.948	ND	
555	17			0.103	1.000	0.944	0.945	264.1	
All above haplotypes ^f	164			0.994	1.000	0.994	1.000	2547.5	
*09	0901	655	284	0.893	0.970	0.798	0.986	29.9	
		645	8	0.025	1.000	0.889	0.886	60.6	
		675	6	0.019	1.000	0.857	0.885	45.5	
		All above haplotypes ^f	298	0.937	0.960	0.754	0.991	23.3	
*10	1001	655	63	1.000	0.890	0.177	1.000	9.1	
*09 & *10	0901/1001	655		0.961	0.996	0.975	0.994	252.8	
*11	1101/1104/1106 ^g	652	72	0.673	0.999	0.973	0.987	881.8	
		642	23	0.215	1.000	0.958	0.969	563.4	
		672	6	0.056	1.000	1.000	0.963	ND	
		All above haplotypes ^f	101	0.944	0.999	0.971	0.998	824.7	
*12	1201	245	31	0.284	1.000	1.000	0.971	ND	
		241	22	0.202	1.000	1.000	0.968	ND	
		351	15	0.138	1.000	1.000	0.965	ND	
		248	13	0.119	1.000	1.000	0.965	ND	
		243	8	0.073	0.999	0.800	0.963	96.1	
		741	7	0.064	1.000	1.000	0.963	ND	
		641	3	0.028	1.000	1.000	0.961	ND	
		All above haplotypes ^f	99	0.908	0.999	0.980	0.996	1189.4	
		1202	345	39	0.534	1.000	1.000	0.987	ND
			365	6	0.082	1.000	1.000	0.975	ND
			All	45	0.616	1.000	1.000	0.990	ND
*13	1301/1302/1307 ^g	2410	76	0.371	1.000	0.987	0.951	935.4	
		2411	43	0.210	1.000	1.000	0.940	ND	
		All above haplotypes ^f	119	0.580	1.000	0.992	0.967	1464.6	
*14	1401	746	51	0.823	0.982	0.510	0.996	44.8	
		656	4	0.065	1.000	1.000	0.979	ND	
		747	3	0.048	1.000	1.000	0.978	ND	
	1402/1403/1404 ^g	458	15	0.625	0.999	0.833	0.997	563.3	
		748	4	0.167	1.000	0.800	0.993	450.7	
		746	47	0.797	0.980	0.470	0.995	40.1	
	1405	746	47	0.797	0.980	0.470	0.995	40.1	
		118	15	0.600	1.000	1.000	0.996	ND	
	1406/1407/1412 ^g	118	15	0.600	1.000	1.000	0.996	ND	
		458	3	0.120	0.994	0.167	0.992	21.6	
*14	All above haplotypes ^f	142	0.8875	1.0000	1.0000	0.9930	ND		
*15	1501	116	177	0.843	0.992	0.894	0.987	101.1	
		356	12	0.057	0.999	0.857	0.927	71.9	
		All above haplotypes ^f	189	0.900	0.991	0.892	0.992	98.5	
	1502	1113	16	0.193	1.000	1.000	0.975	ND	
*16	1602	647	29	0.935	1.000	1.000	0.999	ND	

This table includes the haplotypes with high predictive value >0.8 for DRs among all haplotypes with statistically significantly associated with the particular DRs (All *P*-values less than 0.01 by Fisher's Exact Probability Test), except for [6-5-5] shared by *0901 and *1001, and haplotypes of subtypes of DR*14.

^aSensitivity.

^bSpecificity.

^cPositive predictive value.

^dNegative predictive value.

^eLikelihood ratio.

^fMeans all above haplotypes of given subtype.

^gSeveral alleles among the same DRs were grouped together when they shared haplotypes.

ND, not available due to calculation divided by 0.

Table 6 The distribution of highly predictable haplotypes in each subtype of DRB1*04 and all DRB1*04 type in Caucasian populations

DRB1 type	Haplotype	Number	SS ^a	SP ^b	PPV ^c	NPV ^d	LR ^e
*0401	673	491	0.660	0.861	0.574	0.900	4.8
	683	208	0.280	0.995	0.945	0.830	61.3
	773	14	0.019	1.000	1.000	0.783	ND
	113	8	0.011	0.990	0.229	0.780	1.0
	693	6	0.008	1.000	0.857	0.781	21.2
	173	4	0.005	1.000	1.000	0.780	ND
*0402	666	17	0.944	0.998	0.708	1.000	452.8
	656	1	0.056	1.000	1.000	0.995	ND
*0403/*0407 ^f	673	36	0.750	0.753	0.042	0.995	3.0
	666	6	0.125	0.995	0.250	0.987	23.1
*0404/*0408 ^f	673	310	0.787	0.817	0.362	0.967	4.3
	653	32	0.081	0.999	0.889	0.892	60.5
	663	18	0.046	0.983	0.257	0.886	2.6
	113	15	0.038	0.993	0.429	0.886	5.7
	683	9	0.023	0.929	0.041	0.878	0.3
	353	4	0.010	0.999	0.667	0.884	15.1
*0405	663	50	0.820	0.994	0.714	0.997	135.8
	673	6	0.098	0.743	0.007	0.978	0.4
	113	4	0.066	0.991	0.114	0.983	7.0
DR4 ^g	673	843	0.666	0.994	0.985	0.832	108.1
	683	220	0.174	1.000	1.000	0.669	ND
	663	70	0.055	1.000	1.000	0.638	ND
	653	36	0.028	1.000	1.000	0.632	ND
	666	24	0.019	1.000	1.000	0.630	ND
	773	14	0.011	1.000	1.000	0.628	ND
	693	7	0.006	1.000	1.000	0.626	ND
	173	4	0.003	1.000	1.000	0.626	4.8
	All above haplotypes ^h	1218	0.963	0.994	0.989	0.978	61.3

^aSensitivity.^bSpecificity.^cPositive predictive value.^dNegative predictive value.^eLikelihood ratio.^fSeveral alleles were grouped together when they shared haplotypes.^gThe *P*-values of each haplotype (673, 683, 663, 653, 666, 773, 693 and 173) for any subtype of DR4 are <0.0001, 4.4×10^{-100} , 4.4×10^{-31} , 3.4×10^{-16} , 5.2×10^{-11} , 1.04×10^{-6} , 0.001 and 0.0197, respectively.^hMeans all above haplotypes of the DR4 type.

ND, not available due to calculation divided by 0.

HLA-DRB genes and pseudogenes that form five major haplotype families, broadly designated as the DR1, DR8, DR51, DR52 and DR53 haplotype family groups.¹⁶ DRB1*09 is clearly a member of the DR53 family containing two functional DRB genes (DRB1 and DRB4) and several pseudogenes, such as ψ DRB7, ψ DRB8, and ψ DRB9. On the other hand, DRB1*10 is a member of the DR1 family group that has only one functional DRB1 locus and two pseudogenes ψ DRB6 and ψ DRB9. DR9 and DR10 have also been shown to have different patterns of linkage disequilibrium with the DQ region.^{10,12-14} We have also previously proposed, based on the sequencing of 3'-UT region of DRB1 alleles, that DRB1*10 allele is quite distant evolutionally from DR53 family (DRB1*04, *07 and *09) but more closely related to the DR1/DR2 haplotype family.¹¹ For this reason, the local haplotype similarity of '6-5-5' within the DR region between DR9 and DR10 cannot be due to ancient family relationships.

This raises the possibility that one or more gene conversion or recombination events may have occurred between ancestral haplotypes in these two haplotype families.

As only three microsatellites have been run in this study, we cannot assess the extent of the similarity between DR9 and DR10 in the class II region. D6S2666 and D6S2665 are located between BTNL2 in the central MHC and HLA-DRA. An examination of our previous microsatellite data in Caucasians suggests that the telomeric extent of this haplotype similarity may extend a few hundred thousand base pairs into the central MHC (data not shown).⁶ As the D6S2446 marker is centromeric to the HLA-DQA gene, this region may not in fact be part of the shared region between the DR9 and DR10 haplotypes. However, we will need to carry out a dense SNP map of these haplotypes in order to define this region of similarity more precisely.

Table 7 The distribution of highly predictable haplotypes in each subtype of DRB1*04 and all DRB1*04 type in Korean populations

DRB1	Haplotype	Number	SS ^a	SP ^b	PPV ^c	NPV ^d	LR ^e
*0401	673	26	0.813	0.919	0.107	0.998	10.0
*0403/0407 ^f	673	57	0.877	0.930	0.234	0.997	12.5
	663	4	0.062	1.000	0.800	0.978	163.9
	653	4	0.062	0.975	0.056	0.977	2.4
*0404/0408 ^f	673	26	0.591	0.919	0.107	0.993	7.3
	653	8	0.182	0.977	0.113	0.986	7.7
	683	7	0.159	0.999	0.636	0.986	106.8
*0405/0445 ^f	113	411	0.888	0.990	0.949	0.977	91.4
	653	35	0.076	0.984	0.493	0.839	4.8
	673	10	0.022	0.897	0.041	0.818	0.2
	115	3	0.006	0.985	0.081	0.829	0.4
*0406	673	119	0.944	0.952	0.488	0.997	19.7
	113	3	0.024	0.835	0.007	0.946	0.1
*0410	653	21	0.808	0.981	0.296	0.998	43.6
	673	3	0.115	0.911	0.012	0.991	1.3
DR4 ^g	113	418	0.552	0.992	0.965	0.852	72.6
	673	241	0.318	0.998	0.988	0.792	209.2
	653	70	0.092	0.999	0.986	0.741	182.3
	683	11	0.015	1.000	1.000	0.725	ND
	663	5	0.007	1.000	1.000	0.724	ND
	All above haplotypes ^h	745	0.984	0.990	0.975	0.994	102.1

^aSensitivity.

^bSpecificity.

^cPositive predictive value.

^dNegative predictive value.

^eLikelihood ratio.

^fSeveral alleles were grouped together when they shared haplotypes.

^gThe *P*-values of each haplotype (113, 673, 653, 683 and 663) for any subtype of DR4 are 1.3×10^{-255} , 8.9×10^{-143} , 5.1×10^{-39} , 7.1×10^{-7} and 0.0016, respectively.

^hMeans all above haplotypes of given subtype.

ND, not available due to calculation divided by 0.

These results are intriguing because the association of RA with DRB1*0901 in the Korean populations carries especially high risk when present as a compound heterozygote with DRB1*0405.¹⁹ This is reminiscent of the high risk of DRB1*0401/0404 compound heterozygotes in Caucasians.^{20,21} Whether DR10 confers similar effects has not been previously addressed, since this allele is so uncommon in Asian populations.^{14,19} However, a preliminary analysis of the currently available data suggest that the DRB1*0405/1001 genotype also confers very high risk for RA in the Korean population with OR = 13.3, 95% CI 1.79–99.2 (22/1149 in cases, 1/698 in controls) (Table 8). Clearly, these data need to be replicated. However, they suggest that the haplotype segment that is shared between DR9 and DR10 may have common functional effects.

Both the DR9 and DR10 associations with RA have been commonly ascribed to the similarity of their DRB1 sequences to the 'shared epitope' alleles.²² These RA associated alleles carry the sequence Q-K/R-R-A-A at positions 70–74 of the DRB1 chain. DRB1*10 has the ⁷⁰R-R-R-A-A⁷⁴ motif in this region, and DRB1*09 encodes a ⁷⁰R-R-R-A-E⁷⁴ motif. Notably, the glutamic acid (E) in position 74 of DRB1*09 is also found in DRB1*0403, an allele that is not associated with RA.²⁰ While the triple arginine sequence at 70–72 may confer functional

similarity to the other shared epitope alleles, it is tempting to speculate that additional allelic variation on DR9 and DR10 haplotypes may at least in part account for the patterns of association with RA. Addressing this hypothesis will require a thorough sequence analysis and comparison of these two haplotypes, as well as more focused population association studies to address their possible interactions with other DRB1 risk alleles.

Materials and methods

Study populations

Caucasians (*n* = 1698) in the present study were 1322 RA patients, who were drawn from the North American Rheumatoid Arthritis Consortium (NARAC) collection²³ and 376 healthy controls in the New York Cancer Project (NYCP) who were cancer-free at the time of enrollment.²⁴ Korean populations (*n* = 1394) consisted of 746 RA patients who were recruited consecutively from the outpatient clinic of The Hospital for Rheumatic Diseases, Hanyang University, Seoul, South Korea and 648 healthy volunteers who were ethnically matched and enrolled from among the staff at the same hospital (nurses, paramedical personnel and laboratory workers).

Table 8 The risk of various combination of HLA-DRB1*0405, *0901, and *1001 for RA compared with control in Korean population

Genotype	RA (n = 1171)	Control (n = 698)	OR	95% CI	P
0405/0405	50	7	4.4	1.99–9.77	7.1×10^{-5}
0405/0901	90	4	14.4	5.28–39.5	1.0×10^{-11}
0405/1001	22	1	13.3	1.79–99.2	0.0009
0901/0901	24	6	2.4	0.98–5.93	0.048
1001/1001	0	0	NA	NA	NA
0901/1001	7	0	NA	NA	0.04

This analysis was performed in patients with RA and controls with HLA typing using high-resolution method in Korean populations. The number of these data is larger than the number of Korean subjects using three-marker microsatellite method in this study. NA, not available.

All RA patients in both groups met the American College of Rheumatology 1987 classification criteria for RA.²⁵ Written informed consent was obtained from all participants.

Laboratory procedures

Caucasian RA and control participants were HLA-DRB1 typed using the SSOP low-resolution method. All individuals with DRB1*04 were subsequently tested using a medium resolution panel, providing subtype results. In all Korean participants, typing for HLA-DRB1 was performed by polymerase chain reaction sequence-based typing (PCR-SBT) method using the reference protocol of the Twelfth International Histocompatibility Workshop.⁵

Three microsatellite markers nearby HLA-DRB1, which were chosen from the 54 markers distributed throughout the HLA complex as described elsewhere,^{6,7} were D6S2666, D6S2665 and D6S2446. The exact marker locations were determined from University of California, Santa Cruz, Genome Bioinformatics Web site, the Web site of the Sanger Institute, and the Web site of dbMHC microsatellite markers.

Fluorescent genotyping was performed in 96-well plates using three fluorescently labeled oligonucleotides amplified in a single multiplex reaction. PCR (40 cycles) was performed on a MJ Research tetrad (10 μ l reactions: 2 ng genomic DNA, 2.5 mM MgCl₂, 0.2 mM dNTPs (Amersham Biosciences, Piscataway, NJ, USA) 0.242 μ mol of D6S2666 5' and 3' primers, 0.161 μ mol of D6S2665 5' and 3' primers, and 0.081 μ mol of D6S2446 5' and 3' primers (Integrated DNA Technologies, Coralville, IA, USA), 0.04 U of Amplitaq Gold DNA Polymerase (Perkin-Elmer, Shelton, CT, USA) in 10 \times PCR Buffer II (Perkin-Elmer, Shelton, CT, USA). Amplified products were electrophoresed with 500 ROX™ size standard (Applied Biosystems, Foster City, CA, USA) through Performance Optimized Polymer-4 (ABI) for 35 min at 15 Volts using a 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Fragment sizes were calculated using Genescan v3.6 analysis and analyzed using Genotyper v3.6NT software (ABI). Each genotype was read independently by two individuals, to ensure accurate calling of alleles. There were 11 samples and 30 samples in which reaction of all three markers failed,

resulting in a total of 1687 (1313 RA patients and 374 controls) and 1364 (744 RA patients and 620 controls) samples for haplotype reconstruction in Caucasians and Koreans, respectively.

Data analysis

To match the particular three-marker haplotype and DRB1, haplotype reconstruction was performed for four markers including three markers and DRB1 allele using PHASE v2.1.1,^{8,9} which employs a Markov Chain Monte Carlo algorithm that is based on the coalescent model. The reconstructed haplotypes were analyzed for associations with the DR types using Fisher's Exact Probability Test. The adjusted logit estimate of the OR and the corresponding CI were computed by Woolf modification. All analyses were conducted using the SAS software package, version 9.1 (SAS Institute, Cary, NC, USA). Sensitivity, specificity, positive predictive value, negative predictive value and likelihood ratio were then calculated in order to evaluate the relationship between the HLA typing and the haplotypes that were a frequency >0.1% within each population and were found to be significantly associated with a particular DR type.

Acknowledgements

This work was supported in part by a grant of the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (01-PJ3-PG6-01GN11-0002). The North American Rheumatoid Arthritis Consortium is supported by the National Institutes of Health (RO1-AR44222 and NO1-AR-2-2263) and by the National Arthritis Foundation. H-S Lee is generously supported by the Eileen Ludwig Greenland Center for Rheumatoid Arthritis at the Feinstein Institute.

References

- 1 Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens* 2004; **64**: 631–649.
- 2 Cooke GS, Hill AV. Genetics of susceptibility to human infectious disease. *Nat Rev Genet* 2001; **2**: 967–977.
- 3 Larsen CE, Alper CA. The genetics of HLA-associated disease. *Curr Opin Immunol* 2004; **16**: 660–667.
- 4 Buckner JH, Nepom GT. Genetics of rheumatoid arthritis: is there a scientific explanation for the human leukocyte antigen association? *Curr Opin Rheumatol* 2002; **14**: 254–259.
- 5 Begnon JD, Fernandez-Vina MA. Protocols of the 12th International Histocompatibility Workshop for typing of HLA class II alleles by DNA amplification by the polymerase chain reaction (PCR) and hybridization with sequence specific oligonucleotide probes (SSOP). In: Charron D (ed). *Genetic Diversity of HLA: Functional and Medical Implication*. EDK: Paris, 1997, pp. 584–595.
- 6 Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X *et al*. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. *Am J Hum Genet* 2002; **71**: 585–594.
- 7 Graham RR, Ortmann WA, Langefeld CD, Jawaheer D, Selby SA, Rodine PR *et al*. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet* 2002; **71**: 543–553.

- 8 Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003; **73**: 1162–1169.
- 9 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.
- 10 Klitz W, Maiers M, Spellman S, Baxter-Lowe LA, Schmeckpeper B, Williams TM *et al*. New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. *Tissue Antigens* 2003; **62**: 296–307.
- 11 Merryman P, Silver J, Gregersen PK, Solomon G, Winchester R. A novel association of DQ alpha and DQ beta genes in the DRw10 haplotype. Determination of a DQw1 specificity by the DQ beta-chain. *J Immunol* 1989; **143**: 2068–2073.
- 12 Song EY, Park MH, Kang SJ, Park HJ, Kim BC, Tokunaga K *et al*. HLA class II allele and haplotype frequencies in Koreans based on 107 families. *Tissue Antigens* 2002; **59**: 475–486.
- 13 Lee KW, Oh DH, Lee C, Yang SY. Allelic and haplotypic diversity of HLA-A, -B, -C, -DRB1, and -DQB1 genes in the Korean population. *Tissue Antigens* 2005; **65**: 437–447.
- 14 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**: 522–529.
- 15 Bergstrom TF, Josefsson A, Erlich HA, Gyllensten U. Recent origin of HLA-DRB1 alleles and implications for human evolution. *Nat Genet* 1998; **18**: 237–242.
- 16 Hohjoh H, Ohashi J, Takasu M, Nishioka T, Ishida T, Tokunaga K. Recent divergence of the HLA-DRB1*04 allelic lineage from the DRB1*0701 lineage after the separation of the human and chimpanzee species. *Immunogenetics* 2003; **54**: 856–861.
- 17 Gaudieri S, Leelayuwat C, Tay GK, Townend DC, Dawkins RL. The major histocompatibility complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. *J Mol Evol* 1997; **45**: 17–23.
- 18 Lai Y, Sun F. The relationship between microsatellite slippage mutation rate and the number of repeat units. *Mol Biol Evol* 2003; **20**: 2123–2131.
- 19 Lee HS, Lee KW, Song GG, Kim HA, Kim SY, Bae SC. Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901. *Arthritis Rheum* 2004; **50**: 3468–3475.
- 20 Hall FC, Weeks DE, Camilleri JP, Williams LA, Amos N, Darke C *et al*. Influence of the HLA-DRB1 locus on susceptibility and severity in rheumatoid arthritis. *QJM* 1996; **89**: 821–829.
- 21 MacGregor A, Ollier W, Thomson W, Jawaheer D, Silman A. HLA-DRB1*0401/0404 genotype and rheumatoid arthritis: increased association in men, young age at onset, and disease severity. *J Rheumatol* 1995; **22**: 1032–1036.
- 22 Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* 2004; **5**: 151–157.
- 23 Gregersen PK. The North American Rheumatoid Arthritis Consortium—bringing genetic analysis to bear on disease susceptibility, severity, and outcome. *Arthritis Care Res* 1998; **11**: 1–2.
- 24 Mitchell MK, Gregersen PK, Johnson S, Parsons R, Vlahov D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J Urban Health* 2004; **81**: 301–310.
- 25 Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; **31**: 315–324.

Supplementary Information accompanies the paper on Genes and Immunity website (<http://www.nature.com/gene>)