M. Malkki*
R. Single*
M. Carrington
G. Thomson
E. Petersdorf

MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies

Key words:
haplotype; hematopoietic cell transplantation; HLA; MHC; microsatellite

Acknowledgments:
Authors thank Dr Mike Cullen for helpful discussions and Sharie Cheng, Mark Gatterman, and Anita Harding for outstanding technical support.

This work was supported by grants CA18029, AI49213 and AI33484 (for MM and EWP), NO1-CO-12400 (for MC), AI49213 (for GT and RS), P30CA22435 (for RS), and GM 35326 (for GT) from the National Institutes of Health (NIH), and DE-FG02-00ER45828 (for RS) from the US Department of Energy.

Abstract: Twenty-two human major histocompatibility complex (MHC) region microsatellite (Msat) markers were studied for diversity and linkage disequilibrium (LD) with HLA loci in hematopoietic cell transplant recipients and their HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 allele-matched unrelated donors. These Msats showed highly significant LD over much of the MHC region. The Msat diversity of five common Caucasian haplotypes (HLA-A1-B8-DR3, A3-B7-DR15, A2-B44-DR4, A29-B44-DR7, and A2-B7-DR15) was examined using a new measure called 'haplotype specific heterozygosity' (HSH). Each of the five haplotypes had at least one Msat marker with an HSH value of zero indicating that only one Msat allele was observed for the particular HLA haplotype. In addition, the ability of Msats to predict HLA-A-B-DRB1 haplotypes was studied. Over 90% prediction probability of two common haplotypes (HLA-A1-B8-DR3 and HLA-A3-B7-DR15) was achieved with information from three Msats (D6S265/D6S2787/D6S2894 and D6S510/D6S2810/D6S2876, respectively). We demonstrate how the HSH index can be used in the selection of informative Msats for transplantation and disease association studies. Markers with low HSH values can be used to predict specific HLA haplotypes or multilocus genotypes to supplement the screening of HLA-matched donors for transplantation. Markers with high HSH values will be most informative in studies investigating MHC region disease-susceptibility genes where HLA haplotypic effects are known to exist.

A high-density map of the human major histocompatibility complex (MHC) region microsatellites (Msats) is now available (1–6). Accurate and reproducible molecular methods have been applied to study di-, tri-, tetra-, and penta-nucleotide repeats and complex repetitions among unrelated individuals (1–4).

Authors' affiliations:

M. Malkki[1]*,
R. Single[2]*,
M. Carrington[3],
G. Thomson[4],
E. Petersdorf[1,5]

[1]The Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[2]The Department of Mathematics and Statistics, University of Vermont, Burlington, VT, USA

[3]Basic Research Program, SAIC Frederick Inc., Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD, USA

[4]Department of Integrative Biology, University of California, Berkley, CA, USA

[5]The Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA

Correspondence to:
Mari Malkki, PhD
Division of Clinical Research, D4-100
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue North
Seattle
WA 98109
USA
Tel.: +1 206 667 5797
Fax: +1 206 667 2917
e-mail: mmalkki@fhcrc.org

Received 29 March 2005, revised 23 May 2005, accepted for publication 25 May 2005

doi: 10.1111/j.1399-0039.2005.00453.x

Tissue Antigens 2005: **66:** 114–124

*These authors contributed equally to this manuscript.

MHC region Msats have long been used for genetic analyses including disease mapping to understand susceptibility to autoimmune and infectious diseases (7–14) and characterize regions of recombination within the MHC (15–18). Use of Msats has recently been proposed as an approach for identifying non-HLA markers within the MHC that could function as transplantation determinants (19, 20) and for the selection of potential donors for transplantation (21–24).

MHC Msats show considerable polymorphism and strong linkage disequilibrium (LD) with the classical HLA loci (9, 23). Several well-defined extended HLA haplotypes (HLA-A1-B8-DR3, HLA-A30-B18-DR3, HLA-A29-B44-DR7, and HLA-A2-B44-DR4) demonstrate strong LD with Msat alleles and conservation of the haplotypes (23). In disease association studies, where samples are enriched for predisposing HLA class I and class II alleles and haplotypes (25) such as HLA B27 (ankylosing spondylitis), DRB1*0301-DQB1*0201 (type I diabetes), DRB1*0401-DQB1*0302 (type I diabetes and rheumatoid arthritis), and DRB1*1501-DQB1*0602 (multiple sclerosis and narcolepsy), certain Msat markers will be more informative than others due to LD. The full extent to which Msats are in LD with different HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 haplotypes is currently unknown.

We studied 31 Msats that were evenly distributed across the MHC region for level of diversity and LD with HLA loci and assessed the degree to which these markers can be used to predict common HLA haplotypes in 338 transplant recipients and their HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 allele-matched donors. We introduce a measure of Msat diversity on specific HLA-A-B-DRB1-defined haplotypes called 'haplotype specific heterozygosity' (HSH) and describe how this statistic may be used in the selection of informative Msat markers for the analysis of HLA haplotypes in studies of disease association and transplantation.

Materials and methods

Study population

The study population consisted of 338 hematopoietic cell transplant recipients and their HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 allele-matched unrelated donors. Transplantation was performed at the Fred Hutchinson Cancer Research Center between 1985 and 2003 (a total of 676 individuals or 1352 chromosomes) (26) and for whom DNA was available. All research samples and data were collected according to IRB-approved human subjects' guidelines and protocols. The self-described patient and donor racial backgrounds were 97 and 78% Caucasian, respectively. Patient and donor data were analyzed separately because individuals were matched for HLA loci and therefore not independent. We report analyses for the patient samples only in order to simplify the presentation because results for the patient and donor samples were very similar.

Msat genotyping

Primers used for Msat genotyping are listed in Table 1 (refer to the NCBI website by using UniSTS identifier http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists). Fourteen markers (D6S276, D6S105, D6S2972, D6S510, D6S265, D6S2812, D6S2811, D6S2810, D6S2787, D6S273, D6S2876, D6S2874, D6S2749, and D6S291) were amplified using the 13th IHWG Msat typing protocol developed by Carrington et al. (http://www.ihwg.org/shared/micros.htm). For the remaining 17 markers, panel B conditions from this protocol were used. Fluorescence labeling of the 5′ primer was as follows: 6-FAM dye for D6S2239, D6S1558, D6S1001, D6S2902, D6S2920, D6S2894, and D6S2883; NED dye for D6S2931, D6S2907, D6S2863, D6S248, and D6S464; TET dye for D6S2891, D6S2792, D6S2222, and D6S2223; and HEX dye for RH71009. All samples were electrophoresed on the ABI Prism® 377 sequencer and analyzed using GENESCAN v3.1.2 and GENOTYPER v2.5 software (ABI). Ten to 50 reference DNAs from IHWG Core Cell and Gene Bank (http://www.ihwg.org/shared/cbankover.htm) were typed for quality control purposes for each marker.

Statistical methods

Allele frequencies were calculated by direct gene (allele) counting. The heterozygosity (gene diversity) statistic, based on expected Hardy–Weinberg proportions (HWP), was computed as $H = 1 - \sum_{i=1}^{k} p_i^2$, where k indicates the number of alleles at a locus and p_i the allele frequency. Observed genotype frequencies were tested against HWP using the exact test of Guo and Thomson (27) as implemented in Arlequin (28). The Bonferroni and Hochberg (29) procedures were used to control the type I error rate for the total number of markers tested. For loci where significant deviation from HWP was found, the direction of deviation (excess homozygotes or heterozygotes) was assessed. The above statistics and the multilocus analyses described below were computed using PyPop (30, 31).

The frequency of marker matching for a given Msat was defined as the percentage of donor/patient pairs that matched both alleles at the genotype level for that marker. This matching frequency was also computed for subsets of the data based on similar HLA genotype (e.g. those with one HLA-B*4402 allele *vs* those with one B*4403 allele) where sample size allowed.

Msats evaluated in the study population

Marker[a]	Alias	UniSTS	Genomic location build 35.1	Size range (bp)	Number of patients typed	Number of donors typed	Number of alleles (K)[b]	K needed for 50% frequency[c]	HWP deviation[c]	Frequency of marker matching
D6S276		256855	24293842	63–151	323	322	18	3		0.11
D6S2239		239126	26181749	104–120	241	241	5	2	P/D	0.30
D6S1558	AFMA192WG9	49653	27140256	245–259	316	309	10	1		0.31
D6S2223		239116	27765822	167–179	253	252	6	1	P/D	0.47
D6S464	AFM323VB5	72314	27839683	202–226	317	317	15	2	P	0.32
RH71009		51123	27853520	174–186	303	315	8	1	P	0.47
D6S105		256856	27879224	144–168	338	338	13	2		0.21
D6S1001		149950	27923949	176–200	317	316	15	2		0.32
D6S2222		239114	28041225	236–262	244	250	7	2	P	0.43
D6S248		147257	28804174	269–287	309	306	18	2		0.22
D6S2972	MOGCA	239102	29753470	118–150	319	323	16	3	P/D	0.53
D6S2863		464142	29831878	179–209	244	236	14	3		0.62
HLA-A			30018310				24	3		1.00
D6S510		148725	30049247	176–204	302	303	11	2		0.73
D6S265		256850	30127439	176–200	323	323	12	2		0.71
D6S2907	C5_3_1	464176	30344189	125–149	243	240	8	2	D	0.52
D6S2902		464245	31203274	172–202	251	247	10	2	P/D	0.65
D6S2812	HLACCA1	256857	31257541	103–125	336	336	12	2		0.75
D6S2931		464254	31312050	234–290	250	248	14	4		0.50
HLA-C			31344508				22	4		1.00
D6S2811	HLABCCA2	256858	31367095	96–138	301	301	19	4		0.66
HLA-B			31429846				49	5		1.00
D6S2810	MIB	256842	31457331	257–291	302	301	18	3		0.68
D6S2792	3-7 TNF A/B	464283	31643332	147–175	243	238	14	4		0.60
D6S2787	BAT2CA	239131	31685347	138–162	302	303	13	3		0.63
D6S273		256846	31791664	139–165	276	322	10	2		0.55
D6S2920	3-3	464318	32153714	200–228	242	244	15	3		0.29
D6S2894	NOTCH4	464327	32299495	325–355	239	241	10	2		0.73
D6S2891	3-3A	464341	32406317	161–191	246	243	12	2	P/D	0.57
D6S2883	DRACA	464360	32511199	241–261	234	228	8	3		0.83
DRB1			32654527				31	4		1.00
DQB1			32735642				15	3		1.00
D6S2876	G51152	256839	32778014	193–245	283	284	15	2		0.85
D6S2874	TAP1CA	156302	32927618	187–209	321	320	9	2		0.44
D6S2749	RING3CA	256859	33050865	221–239	322	323	9	2		0.26
DPB1			33151738				24	2		0.23
D6S291		153527	36373526	164–186	320	317	9	2		0.21

[a]HLA loci are included for the purpose of orientation with respect to the Msats studied.
[b]Patient data were used.
[c]Patient (P), donor (D), blank indicates no deviation from HWP.

Table 1

Haplotype frequencies were estimated via the expectation-maximization algorithm and used to compute two measures of overall LD. LD was measured using Hedrick's multiallelic D' statistic (32) and W_n, also known as Cramer'sV Statistic (33). In the formulae below, p_i and q_j are allele frequencies at two loci with I and J alleles, respectively, $D'_{ij} = D_{ij}/D_{max}$, $D_{ij} = h_{ij} - p_i q_j$, and h_{ij} are the estimated haplotype frequencies.

$$D' = \sum_{i=1}^{I}\sum_{j=1}^{J} p_i q_j \left| D'_{ij} \right|$$

$$W_{\mathrm{n}} = \left[\frac{\sum_{i=1}^{I}\sum_{j=1}^{J} D_{ij}^2 / p_i q_j}{\min(I-1, J-1)} \right]^{1/2}$$

The significance of the LD between pairs of loci was tested using a permutation test with 1500 permutations for each locus pair. Because each locus was tested with each of the other loci, a Bonferroni correction for one less than the total number of loci was used for each P-value in order to correct for the number of comparisons in a given row of the matrix of LD values.

In addition to the overall heterozygosity for each Msat, we computed the HSH for each Msat associated with the five most frequent Caucasian HLA-A-B-DRB1 haplotypes. Although data for the HLA-C and HLA-DQB1 were available, we chose to focus on common HLA-A-B-DRB1 haplotypes in order to have a sufficient number of individual haplotypes. The HSH is the heterozygosity of a particular Msat given a specific HLA haplotype. It is computed separately for each HLA haplotype by normalizing the Msat allele frequencies found on the specific HLA haplotype and then calculating the above heterozygosity statistic using the normalized frequencies. The normalized frequencies for these haplotype specific Msat alleles are $p_i = h_i / \sum_{j=1}^{k} h_j$and then $HSH = 1 - \sum_{i=1}^{k} p_i^2$, where k is the number of Msat alleles observed on the specific HLA-A-B-DRB1 haplotype, and $h_1, \ldots, h_k$ are the frequencies of the four-locus Msat – HLA-A-B-DRB1 haplotypes.

For the prediction of specific HLA haplotypes by Msat alleles, frequencies for haplotypes consisting of HLA-A, HLA-B, HLA-DRB1, and one or more Msats were first estimated. Then, probabilities of different HLA-A-B-DR haplotypes, given a specific Msat allele, were computed by summing up the appropriate haplotype frequencies. The sensitivity is defined as the probability of observing the Msat allele(s) given that the particular HLA haplotype was observed. The specificity is the probability of not observing the Msat allele(s) given that the HLA haplotype was not observed. The positive predictive value (PPV) is defined as the probability of observing the HLA haplotype given that the specific Msat allele(s) was observed. The negative predictive value is the probability of not observing the HLA haplotype given that the specific Msat allele(s) was not observed. Higher values for each of these statistics indicate, in slightly different ways, that there is a stronger association of the Msat allele(s) with the HLA haplotype.

Results

Allele frequencies and HWP of Msats

The highest level of polymorphism, measured by the number of alleles, was observed for D6S2811 (19 alleles) and the lowest with the D6S2239 (five alleles) marker (Table 1). The frequency of Msat alleles was not significantly different between recipient and donor samples. At each Msat locus one to four major alleles represented 50% of the total frequency.

Deviation from HWP was tested for each Msat marker and HLA locus in the Caucasian samples. None of the HLA loci showed a significant deviation from HWP. After correcting for multiple comparisons, D6S2239, D6S2223, D6S2972, D6S2902, and D6S2891 displayed significant deviation from HWP in both donor and patient samples. In addition, significant deviation from HWP was observed for D6S464, RH71009, and D6S2222 in patients and D6S2907 in the donors. We investigated the contributions of individual homozygotes and heterozygotes to the overall result for the above markers; several deviations were in the direction of excess homozygotes, suggesting non-amplification of certain alleles. Accordingly, these nine markers were removed from further data analysis, leaving 22 markers in the analyses.

LD between Msats and HLA

Patterns of overall LD, measured as W_n and D', in the patient group are shown in Fig. 1. Both W_n and D' are standardized measures that range from zero to one, with higher values indicating stronger LD. While the two statistics are correlated, they are influenced differently by various aspects of the strength of LD, such as sensitivity to the number of alleles or estimation of low-frequency haplotypes. The strongest LD was observed between HLA genes HLA-C-B ($W_n = 0.81$, $D' = 0.96$) and HLA-DRB1-DQB1 ($W_n = 0.83$, $D' = 0.95$). Msat markers had the highest degree of LD with the nearest HLA locus. The strongest associations were observed for HLA-A-D6S265 ($W_n = 0.81$, $D' = 0.86$), HLA-A-D6S510 ($W_n = 0.67$, $D' = 0.87$), HLA-B-D6S2810 ($W_n = 0.71$, $D' = 0.89$), HLA-B-D6S2811 ($W_n = 0.68$, $D' = 0.85$), HLA-DRB1-D6S2883 ($W_n = 0.73$, $D' = 0.82$), HLA-DRB1-D6S2876 ($W_n = 0.68$, $D' = 0.90$), and HLA-DQB1-D6S2876 ($W_n = 0.63$, $D' = 0.81$). Each of these associations was highly significant. These results demonstrate that the MHC region

Fig. 1. **Matrix of overall linkage disequilibrium (LD) values for each pairwise combination of markers in 338 transplant recipients.** Loci are listed in order from the most telomeric marker D6S276 to the most centromeric marker D6S291. Results for the D′ and W_n measures are in the lower and upper triangles of the matrix, respectively. Both W_n and D′ are standardized measures that range from zero to one as shown by the colored squares with darker shading indicating stronger LD.

Msat markers in this study show strong LD with the classical HLA loci.

Donor-recipient Msat matching

Among the HLA-matched transplant pairs, the frequency of donor-recipient Msat matching ranged from 11% for the most telomeric marker (D6S276), to over 80% for D6S2883 and D6S2876 (Table 1). Not surprisingly, there was a strong negative correlation between the frequency of donor-recipient Msat matching and the distance to the nearest classical HLA locus (correlation $= -0.73$, P-value < 0.001); the negative value indicates that the frequency of Msat matching increased as the distance between Msat and HLA loci decreased. Apart from marker D6S2920, all Msats between D6S2863 and D6S2876 had a matching frequency of more than 50%, while other markers outside this region had a matching frequency of less than 50%. These results indicate that HLA identical unrelated individuals have differences in their Msat markers.

HSH

The HSH provides a summary of the distribution of Msat alleles on HLA-defined haplotypes and gives additional haplotype- specific information about the diversity of the Msats and how this diversity

varies from one HLA-defined haplotype to another. Diversity of Msats among the five most frequent Caucasian haplotypes in the patient group is displayed in Table 2. The table contains the overall heterozygosity (gene diversity) index and HLA-A-B-DRB1 HSH for each Msat marker. Msat markers closest to the HLA classical loci showed a trend for increased overall heterozygosity. This heterozygosity was negatively correlated (correlation = −0.30) with the distance to the nearest HLA locus, but this correlation was not statistically significant. The overall heterozygosity ranged from 0.57 (D6S1558) to 0.90 (D6S2811 and D6S2931) with the highest values at markers near the HLA-B locus. As expected, the HSH value was lower than the overall heterozygosity value in all instances, and these two statistics were similar for markers most distant from a classical HLA locus. The haplotype-specific values were substantially lower for certain HLA haplotypes. For example, for D6S2811 and D6S2931, the HSH values dropped below 20% on

Marker heterozygosity and haplotype-specific heterozygosity in the patient population

		HLA-A-B-DRB1 haplotype-specific Msat heterozygosity (number of alleles)				
Marker[a]	Marker Hetz[b]	0101-0801-0301	0301-0702-1501	0201-4402-0401	2902-4403-0701	0201-0702-1501
D6S276	0.83	0.70 (8)	0.57 (4)	0.70 (5)	0.79 (6)	0.77 (6)
D6S1558	0.57	0.12 (2)	0.18 (2)	0.33 (3)	0.40 (3)	0.45 (4)
D6S105	0.81	0.34 (6)	0.49 (3)	0.31 (3)	0.50 (3)	0.74 (5)
D6S1001	0.73	0.54 (6)	0.37 (3)	0.33 (3)	0.49 (2)	0.60 (4)
D6S248	0.84	0.37 (6)	0.39 (2)	0.63 (4)	0.54 (3)	0.75 (5)
D6S2863	0.83	0.04 (2)	0.08 (2)	0.46 (3)	0.15 (2)	0.48 (2)
HLA-A						
D6S510	0.79	0.03 (2)	0 (1)	0.14 (2)	0 (1)	0.49 (3)
D6S265	0.78	0 (1)	0.06 (2)	0.16 (2)	0.10 (2)	0.39 (3)
D6S2812	0.78	0.32 (5)	0.11 (2)	0 (1)	0 (1)	0 (1)
D6S2931	0.90	0.17 (2)	0.77 (6)	0.45 (4)	0.37 (4)	0.34 (3)
HLA-C						
D6S2811	0.90	0.15 (4)	0.12 (2)	0.39 (3)	0 (1)	0.21 (2)
HLA-B						
D6S2810	0.88	0.30 (4)	0.06 (2)	0.23 (3)	0 (1)	0.21 (2)
D6S2792	0.88	0.10 (3)	0.07 (2)	0.43 (3)	0.44 (2)	0.27 (3)
D6S2787	0.80	0.09 (3)	0.17 (3)	0.14 (2)	0.24 (3)	0.12 (2)
D6S273	0.80	0.22 (5)	0.13 (3)	0 (1)	0.13 (2)	0.22 (2)
D6S2920	0.86	0.34 (5)	0.44 (4)	0 (1)	0.65 (3)	0.75 (5)
D6S2894	0.74	0.07 (2)	0.07 (2)	0 (1)	0 (1)	0 (1)
D6S2883	0.85	0.12 (2)	0.19 (3)	0.10 (2)	0 (1)	0 (1)
HLA-DRB1						
HLA-DQB1						
D6S2876	0.80	0.15 (5)	0.07 (2)	0 (1)	0.12 (2)	0.41 (5)
D6S2874	0.62	0.22 (3)	0.12 (3)	0.06 (2)	0 (1)	0.59 (5)
D6S2749	0.79	0.50 (5)	0.60 (5)	0.45 (4)	0.62 (4)	0.64 (3)
HLA-DPB1						
D6S291	0.72	0.61 (6)	0.57 (4)	0.53 (4)	0.42 (3)	0.45 (3)
Frequency of HLA-A-B-DRB1 haplotype	–	0.112	0.053	0.052	0.028	0.026
Number of haplotypes[c]	–	76	36	35	19	18

[a]HLA loci are included for the purpose of orientation with respect to the Msats studied.
[b]Overall marker heterozygosity.
[c]Fractional values, a consequence of the EM estimation of haplotype frequencies, have been rounded.

Table 2

the HLA-A*0101-B*0801-DRB1*0301 (A1-B8-DR3) haplotype indicating reduced Msat variability. Each of the five HLA-A-B-DR haplotypes described in Table 2 had at least one Msat with an HSH value of zero indicating that only one Msat allele was observed at that Msat locus for the specific HLA haplotype. The majority of Msat markers that did not have at least one zero value are located telomeric of HLA-A.

Prediction of HLA-A-B-DRB1 haplotype by Msats

The utility of Msats to predict specific HLA-A-B-DRB1 haplotypes was assessed. We chose two sets of three markers to demonstrate how Msat HSH and LD values can be used to select markers for predicting HLA-A-B-DRB1 haplotypes. We focused on the two most common HLA haplotypes A*0101-B*0801-DRB1*0301 (A1-B8-DR3) and A*0301-B*0702-DRB1*1501 (A3-B7-DR15), which had frequencies of 11.2 and 5.3%, respectively, in the patient group (Table 2). We considered Msat markers that were located near HLA-A, HLA-B, and HLA-DRB1, had low HSH values and high levels of LD with the nearest HLA locus. Marker D6S265, which had no variability (HSH = 0), along with D6S2787 and D6S2894, which had HSH values below 10%, were selected for A1-B8-DR3 haplotypes. Markers D6S510, D6S2810, and D6S2876 were selected for A3-B7-DR15 haplotypes because they had HSH values below 10%.

The ability of these Msat markers to predict A1-B8-DR3 and A3-B7-DR15 haplotypes for the patient population is presented in Table 3. The D6S265*180 allele with A1-B8-DR3 and D6S510*192 allele with A3-B7-DR15 had a sensitivity level of 1.0. This corresponds to the HSH value of zero indicating no Msat variability. The lower specificity values on these haplotypes, 0.76 for D6S265 and 0.81 for D6S510, show that the D6S265*180 and D6S510*192 alleles do occur on other HLA haplotypes (roughly 24 and 19% of the time, respectively). The more relevant statistic for the prediction of specific HLA haplotypes is the PPV, that is the probability of observing the particular HLA haplotype in the presence of the particular Msat allele. All but three of the PPV results for the individual markers in Table 3 were below 50%. However, predictions using three markers together were substantially higher, reaching over 90% in some cases. These results demonstrate that it is necessary to consider different measures of the affinity of Msats alleles for HLA haplotypes depending on the goal of the study. The results for the donor group were very similar to those presented for the patient group shown in Tables 2 and 3.

Prediction of HLA haplotypes by Msats in the patient population

A-B-DRB1 haplotype	Msat name(s)	Msat allele(s)	Sens[a]	Spec[b]	PPV[c]	NPV[d]	Number[e]
0101-0801-0301	D6S265	180	1	0.76	0.36	1	76
0101-0801-0301	D6S2787	140	0.95	0.95	0.67	0.99	61
0101-0801-0301	D6S2894	352	0.96	0.95	0.72	1	53
0101-0801-0301	D6S265-D6S2787-D6S2894	180-140-352	0.94	1	0.96	0.99	47
0101-0801-0301	D6S510	178	0.98	0.78	0.34	1	62
0101-0801-0301	D6S2810	281	0.83	0.89	0.46	0.98	53
0101-0801-0301	D6S2876	217	0.92	0.92	0.59	0.99	59
0101-0801-0301	D6S510-D6S2810-D6S2876	178-281-217	0.75	0.99	0.93	0.97	46
0301-0702-1501	D6S265	176	0.97	0.91	0.37	1	31
0301-0702-1501	D6S2787	144	0.91	0.77	0.19	0.99	31
0301-0702-1501	D6S2894	346	0.96	0.62	0.13	1	25
0301-0702-1501	D6S265-D6S2787-D6S2894	176-144-346	0.92	0.99	0.88	1	22
0301-0702-1501	D6S510	192	1	0.81	0.22	1	32
0301-0702-1501	D6S2810	267	0.97	0.82	0.23	1	31
0301-0702-1501	D6S2876	223	0.97	0.85	0.26	1	28
0301-0702-1501	D6S510-D6S2810-D6S2876	192-267-223	0.9	0.99	0.79	0.99	27

[a]Sensitivity; The probability of observing the particular Msat allele given the presence of the particular HLA haplotype.
[b]Specificity; The probability of not observing the particular Msat allele given the absence of the particular HLA haplotype.
[c]Positive predictive value; The probability of observing the particular HLA haplotype given the presence of the particular Msat allele.
[d]Negative predictive value; The probability of not observing the particular HLA haplotype given the absence of the particular Msat allele.
[e]Number; The number of four- or six-locus haplotypes. Fractional values, a consequence of the EM estimation of haplotype frequencies, have been rounded.

Table 3

Discussion

The MHC is the most polymorphic multigene region of the human genome and has been the subject of intense study due to the association of HLA genes with disease susceptibility and its role in transplantation. The MHC is characterized by strong-positive LD between HLA loci, which is apparent at the phenotype (antigen) as well as genotype (allele) level. Msats have greatly accelerated understanding of HLA genomic structure. In this study, we sought to define the extent of Msat diversity and LD of Msat alleles in association with common HLA haplotypes. We chose to study a large population of hematopoietic cell-transplant recipients and their HLA-matched unrelated donors because of the availability of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 allele-level typing data. Because our study included patients who successfully identified HLA-matched donors for transplantation, the population was enriched for common HLA alleles and antigens. Nevertheless, we found a similar range of Msat alleles in our patient and donor population as that described in the literature (1, 6, 9) and in the electronic database 'dbMHC Microsatellite Markers' at NCBI (http://www.ncbi.nlm.nih.gov/mhc). Our study confirms and extends observations from other analyses of HLA region Msats with respect to diversity and LD.

Investigators who have employed Msats for the analysis of disease association have long recognized the technical issues that can arise from polymerase chain reaction (PCR) stutters or slippage (34). Furthermore, new algorithms have been developed to accommodate the technical shortcomings of the Msat genotyping protocols (35). In the current study, nine markers were found to deviate significantly from HWP in Caucasian patient or donor samples. Of these nine, D6S2902, RH71009, D6S2972, D6S464, and D6S2239 had an excess of homozygotes for one or more of the shorter alleles compared with their expectations under HWP. Genotyper analysis of D6S464 and D6S2239 showed several stutter allele peaks [likely caused by strand slippage during PCR amplification (34)], in some samples, which could possibly lead to incorrect genotype calls. Re-evaluation of raw electrophoresis and Genotyper analysis data indicated no obvious technical problems for the remaining markers. We interpret these findings as potentially indicative of misidentification of heterozygotes because of unamplified alleles (36). These results demonstrate the utility of HWP to uncover potential technical issues associated with Msat genotyping.

Another challenge in the interpretation of Msat data across different studies is the lack of standardized nomenclature for comparison of Msat allele fragment sizes that arise from the use of different technical protocols and instrumentation. To this end, the 13th International Histocompatibility Working group has created a quality control panel to create a dictionary of Msat alleles generated using three different technical protocols (37). This effort will provide the necessary platform for development of a nomenclature, which can be used to interpret Msat genotype data across studies (6), in a manner similar to that used for standardization of HLA nomenclature (38).

Permutation tests applied to our study population showed highly significant LD over much of the MHC region (results not shown). Only the most distant markers had nonsignificant LD (primarily marker pairs that included D6S276, D6S1558, D6S291, or D6S2874). In particular, three regions of high LD measured by D' and W_n coincide with classical HLA loci, and the results are consistent with the previously reported recombination hotspot located near D6S2874 (Fig. 1) (39, 40).

We observed a negative correlation between the overall marker heterozygosity and the nearest HLA locus, although this difference was not statistically significant. Slatkin has also demonstrated that balancing selection would increase diversity for a closely linked polymorphic marker, whereas background selection of deleterious mutations and hitchhiking events with advantageous alleles are expected to result in reduced variation at a closely linked neutral marker (41). In our data, Msat markers that were closer to classical HLA genes had higher variability, as measured by heterozygosity. While our markers and samples were not chosen to study this effect, the result is interesting and merits further investigation.

In a recent study, the extended haplotype homozygosity (EHH) was defined in order to examine signatures of recent positive selection using a collection of single nucleotide polymorphisms (SNPs) (42). A stepwise approach was used to study the breakdown of LD over increasing distances from a "core" region by including an increasing number of SNPs at each step. The haplotype homozygosity is computed for each of the core SNP-defined haplotypes. At each step of the EHH computation, additional SNPs are considered with the effect of smoothing the results compared with considering the markers individually. While this may be a good approach for neighboring SNPs, it is not practical for widely spaced polymorphic Msats such as those in our study.

One unmet need of the Msat approach is the lack of a measurement for Msat diversity for different HLA-defined haplotypes. Availability of an easily interpreted statistic alongside LD measurements would provide investigators with an additional tool for assessing new disease-susceptibility genes associated with the at-risk HLA haplotype or discovery of new transplantation determinants. The HSH statistic is computed as a point estimate for each polymorphic marker individually for a given haplotype. A marker in strong LD with an HLA locus and with a low value of HSH indicates that an Msat may be useful in predicting a specific HLA haplotype

or multilocus genotype. Markers in moderate or weak LD with HLA loci or with higher HSH will not be useful in screening for common HLA haplotypes but may be useful for the purpose of identifying regions in the MHC that may be relevant for donor matching in transplantation and in investigating potential disease susceptibility genes in addition to known HLA effects.

Application of the HSH statistic measurement to the analysis of haplotypes in HLA-associated diseases such as type I diabetes, multiple sclerosis, and narcolepsy may provide additional information on effects contributed by non-HLA genes. In a dataset that is enriched for certain HLA haplotypes, for example predisposing HLA-DRB1*0301-DQB1*0201 and HLA-DRB1*0401-DQB1*0302 in type I diabetes and HLA-DRB1*1501-DQB1*0602 in multiple sclerosis and narcolepsy, markers with higher HSH values for these haplotypes would provide more information about potential non-HLA effects. In this situation, a case-control study may match cases and controls for known disease-risk HLA genes in order to minimize the effects of LD between these and the markers under study (43). Also, markers with high HSH among enriched haplotypes would be preferable for use in a homozygous-parent TDT analysis, because only families heterozygous at the Msat marker are informative for this test (44).

The choice of markers as well as the number of informative markers is often a dilemma in designing an approach for disease association studies. It often is necessary to consider different measures of the affinity of Msat alleles for HLA-A-B-DRB1 haplotypes depending on the context of the study. Table 3 shows that a single Msat may have high sensitivity but low PPV. Increasing the number of Msat markers used to predict HLA haplotypes can increase the predictive value but usually at the expense of lowered sensitivity. In the current study, the overall sensitivity of a given panel of markers collectively was roughly equal to the product of the sensitivity for each individual marker. The marker trios in Table 3 were chosen in order to demonstrate their use in this study. As our study population was restricted in the diversity of HLA haplotypes, we did not exhaustively search for the combination of Msat markers that would yield the best predictions for specific HLA-A-B-DRB1 haplotypes.

As the HSH statistic is simple to calculate, it can be computed from some available data from other studies. For example, approximate HSH values can be computed for several risk haplotypes for systemic lupus erythematosus using supplementary data from Graham et al. 2002 (13). These include DRB1*1501-DQB1*0602, DRB1*0301-DQB1*0201, and DRB1*0801-DQB1*0402 haplotypes. In this study, the authors were able to predict specific DR-DQ haplotypes using D6S2666, D6S2665, and D6S2446 markers. The low HSH values for these markers on the DR-DQ haplotypes would have identified them as ideal candidates for this purpose had HSH values been available for them either from pilot data or another study. The authors mention that it is difficult to assign risk to genes on the DRB1*0301 : DQB1*0201 haplotype due to the high level of disequilibrium. HSH values computed for the DRB1*0301-DQB1*0201 haplotypes show considerable variation for adjacent markers. Choosing markers with the highest HSH within a region of interest could help restrict the number of potential Msats for future study in a stratified analysis (45) making it possible to correct for fewer multiple comparisons or reducing possible type-1 errors.

Our study population of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 allele-matched unrelated transplant pairs displayed a wide range of matching for Msat alleles across the MHC Classes I, III, and II regions. The frequency of matching decreased as the distance between Msat markers and the classical HLA loci increased, demonstrating the strength of LD between Msat and HLA loci. In a recent publication describing 100 HLA-matched unrelated Japanese transplant pairs (20), five markers (D6S273, D6S2810, D6S2907, D6S265, and D6S510) of 13 studied overlapped with our Msat panel. The frequency of donor – recipient matching for D6S273, D6S2810, D6S265, and D6S510 (64, 89, 81, and 97%, respectively), was higher than that in our study population (55, 68, 71, and 73%, respectively). This result may be explained in part by the differences in the distribution of Msat alleles, HLA haplotypes, or LD in the two different ethnic groups. Because population stratification could influence the degree of marker matching, we computed the frequency of marker matching in our data both with and without including the non-Caucasian individuals. The maximum difference between the marker matching results for the full data *vs* the Caucasian only data was 3.4%, indicating that the inclusion of the non-Caucasian individuals did not have a large effect on these results.

It has been suggested that Msat information can be used to supplement the selection of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1-matched unrelated donors for transplantation because Msats can provide additional genetic information across the MHC. To further assess the clinical utility of Msats in transplantation, we determined the extent of Msat matching among HLA serologically matched pairs who differed for two allele variants of the same antigen. Because donor-recipient disparity for B*4402 and B*4403 has been associated with graft rejection (46) and acute graft-*vs*-host disease (47), we sought to determine Msat matching for patients and donors who were positive for at least one B*4403 ($n = 33$) or one B*4402 ($n = 75$) allele. The matching frequency for D6S2811 was higher in patients with the B*4403 allele than in patients with the B*4402 allele (83 *vs* 46%, respectively). A similar trend was observed for the D6S2810 marker (79 *vs* 52%, respectively). In patients heterozygous for a B44 haplotype and a non-B44

haplotype, the Msat carried on the non-B44 haplotype could have contributed to the skewed match frequency. However, we found six markers (D6S265, D6S2812, D6S2931, D6S2811, D6S2810, and D6S2787) flanking the HLA-B locus at which a higher percentage of B*4403-positive individuals were matched compared with B44-negative individuals. Likewise, a lower percentage of B*4402-positive individuals were matched compared with B44-negative individuals. These results are in agreement with those from Ahmad et al. (48) who found higher haplotype-specific LD for "haplotopes" defined by HLA-B*4403 *vs* HLA-B*4402. Although we confined our analysis to the B*4402 and 4403 subtypes of B44, these results suggest that Msats are potentially informative for identifying functional markers in LD with known HLA determinants.

Future studies are needed to determine the extent to which results presented here might differ in ethnically diverse populations and when haplotypes are known from segregation analysis rather than inferred. While common haplotypes are estimated accurately (49, 50), lower frequency haplotypes can be less reliable. In addition, our future interest is to study MHC region non-HLA SNPs and compare results with those from Msats to further understand putative functional determinants relevant to transplantation.

It is well known that the success of hematopoietic cell transplantation is directly related to the extent of donor-recipient matching for HLA. Although Msat matching will not replace HLA matching, the potential utility of Msats as adjunctive tool to further assess outcomes is an important area of investigation. The use of Msats as a screening tool for HLA haplotypes is motivated in part by the low cost and efficiency of Msat typing. Ultimately, the clinical utility for donor selection in transplantation will be determined by their ability to identify new MHC resident markers that have clinical significance. Analysis of HLA–Msat associations and use of the HSH can provide information for the design and implementation of a study. These measures can aid investigators in selecting optimal markers for genotyping depending on whether the goal is to broaden coverage of all HLA haplotypes or disease-specific haplotypes or to identify functional markers within the MHC region.

References

1. Martin MP, Harding A, Chadwick R et al. Characterization of 12 microsatellite loci of the human MHC in a panel of reference cell lines. *Immunogenetics* 1998: **47**: 131–8.
2. Tamiya G, Shiina T, Oka A et al. New polymorphic microsatellite markers in the human MHC class I region. *Tissue Antigens* 1999: **54**: 221–8.
3. Matzukawa Y, Makino S, Nakajima K et al. New polymorphic microsatellite markers in the human MHC class II region. *Tissue Antigens* 2000: **56**: 492–500.
4. Matzukawa Y, Makino S, Nakajima K et al. New polymorphic microsatellite markers in the human MHC class III region. *Tissue Antigens* 2001: **57**: 397–404.
5. Cullen M, Malasky M, Harding A, Carrington M. High-density map of short tandem repeats across the human major histocompatibility complex. *Immunogenet* 2003: **54**: 900–10.
6. Gourraud PA, Mano S, Barnetche T, Carrington M, Inoko H, Cambon-Thomsen A. Integration of microsatellite characteristics in the MHC region: a literature and sequence based analysis. *Tissue Antigens* 2004: **64**: 543–55.
7. Carrington M, Marti D, Wade J et al. Microsatellite markers in complex disease: mapping disease associated regions within the human major histocompatibility complex. In: Goldstein DB, Schlötterer C, eds. *Microsatellites: Evolution and Applications*. Oxford, England: Oxford University Press, 1999, 225–37.
8. Price P, Witt C, Allcock R et al. The genetic basis for the association of the 8.1 ancestral haplotype (A1 B8 DR3) with multiple immunological diseases. *Immunol Rev* 1999: **167**: 257–74.
9. Foissac A, Salhi M, Cambon-Thomsen A. Microsatellites in the HLA region: 1999 update. *Tissue Antigens* 2000: **55**: 477–509.
10. Ota M, Mizuki N, Katsuyama Y et al. The critical region for Bechet 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.
11. Ota M, Katsuyama Y, Kimura A et al. A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. *Genomics* 2001: **71**: 263–70.
12. Oka A, Tamiya G, Tomikawa M et al. Association analysis using refined microsatellite markers localizes a susceptibility locus for psoriasis vulgaris within a 111 kb segment telomeric to the HLA-C gene. *Hum Mol Genet* 1999: **8**: 2165–70.
13. Graham RR, Ortman WA, Langefeld CD et al. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet* 2002: **71**: 543–53.
14. Jawaheer D, Li W, Graham R et al. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. *Am J Hum Genet* 2002: **71**: 585–94.
15. Martin M, Mann D, Carrington M. Recombination rates across the HLA complex use of microsatellites as a rapid screen for recombinant chromosomes. *Hum Mol Genet* 1995: **4**: 423–8.
16. Cullen M, Noble J, Erlich H et al. Characterization of recombination in the HLA class II region. *Am J Hum Genet* 1997: **60**: 397–407.
17. Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am J Hum Genet* 2002: **71**: 759–76.
18. Jeffreys AJ, Murray J, Neumann R. High-resolution mapping of cross-overs in human sperm defines a minisatellite associated recombination hotspot. *Mol Cell* 1998: **2**: 267–73.
19. Hanifi Moghaddam P, Zwinderman A, Kazemi M et al. D6STNFa microsatellite locus correlates with CTLp frequency in unrelated bone marrow donor-recipient pairs. *Hum Immunol* 1998: **59**: 295–301.

20. Li S, Kawata H, Katsuyama Y et al. Association of polymorphic MHC microsatellites with GVHD survival and leukemia relapse in unrelated hematopoietic stem cell transplant donor/recipient pairs matched at five HLA loci. *Tissue Antigens* 2004: **63**: 362–8.
21. Carrington M, Wade J. Selection of transplant donors based on MHC microsatellite data. *Hum Immunol* 1996: **50**: 151–4 [Erratum Hum Immunol 1996: 51 106–9].
22. Foissac A, Fort ML, Giraldo P, Abbal M, Raffoux C, Cambon-Thomsen A. Microsatellites in the HLA region potential applications in bone marrow transplantation. *Transplant Proc* 1997: **29**: 2374–5.
23. Foissac A, Fort M, Clayton J et al. Microsatellites in the HLA region HLA prediction and strategies for bone marrow donor registries. *Transplant Proc* 2001: **33**: 491–2.
24. MADO European Consortium 2002, http://www.euromado.org/.
25. Rhodes DA, Trowsdale J. Genetics and molecular genetics of the MHC. *Rev Immunogenet* 1999: **1**: 21–31.
26. Petersdorf EW, Anasetti C, Martin PJ et al. Limits of HLA mismatching in unrelated hematopoietic cell transplantation. *Blood* 2004: **104**: 2976–80.
27. Guo SW, Thompson EA. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 1992: **48**: 361–72.
28. Schneider S, Roessli D, Excoffier L. *Arlequin ver.2000. A Software for Population Genetics Data Analysis.* Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva, Switzerland. (retrieved from http://anthro.unige.ch/arlequin).
29. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988: **75**: 800–2.
30. Lancaster A, Nelson MP, Meyer D, Thomson G, Single RM. PyPop, a software framework for population genomics analyzing large-scale multi-locus genotype data. *Pacific Symposium on Biocomputing* 2003: **8**: 514–25.
31. Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G. Software framework for the Biostatistics Core. In: Hansen JA, Dupont B, eds. *HLA 2004 Immunobiology of the Human MHC.* Seattle, USA: IHWG Press, 2005. (in press)
32. Hedrick PW. Gametic disequilibrium measures proceed with caution. *Genetics* 1987: **117**: 331–41.
33. Cramer H. *Mathematical Methods of Statistics.* Princeton, NJ, Princeton: University Press, 1946.
34. Rodriguez S, Visedo G, Zapata C. Detection of errors in dinucleotide repeat typing by nondenaturing electrophoresis. *Electrophoresis* 2001: **22**: 2656–64.
35. Matsumoto T, Yukawa W, Nozaki Y et al. Novel algorithm for automated genotyping microsatellites. *Nucleic Acids Res* 2004: **32**: 6069–77.
36. Beaumont MA, Bruford MW. Microsatellites in conservation genetics. In: Goldstein DB, Schlötterer C, eds. *Microsatellites: Evolution and Applications.* Oxford, England: Oxford University Press, 1999, 163–82.
37. Mickelson E, Damodaran A, He P et al. Standardization of microsatellite data comparative analysis of results from 13th IHWS participating laboratories and establishment of a microsatellite reference panel and standard nomenclature. In: Hansen JA, Dupont B, eds. *HLA 2004 Immunobiology of the Human MHC.* Seattle, USA: IHWG Press, 2005 (in press).
38. Robinson J, Waller MJ, Parham P et al. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 2003: **31**: 311–4.
39. Cullen M, Erlich H, Klitz W, Carrington M. Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus. *Am J Hum Genet* 1995: **56**: 1350–8.
40. Jeffreys AJ, Ritchey A, Neumann R. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum Mol Genet* 2000: **9**: 725–33.
41. Slatkin M. Hitchhiking and associative overdominance at a microsatellite locus. *Mol Biol Evol* 1995: **12**: 473–80.
42. Sabeti PC, Reich DE, Higgins JM et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002: **419**: 832–7.
43. Hanifi Moghaddam P, de Knijf P, Roep BO et al. Genetic structure of IDDM1 two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Belgian Diabetes Registry. *Diabetes* 1998: **47**: 263–9.
44. Lie BA, Todd JA, Pociot F et al. The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am J Hum Genet* 1999: **64**: 793–800.
45. Valdes AM, Thomson G. Detecting disease-predisposing variants: the haplotype method. *Am J Hum Genet* 1997: **60**: 703–16.
46. Fleischhauer K, Kernan NA, O'Reilly RJ, Dupont B, Yang SY. Bone marrow-allograft rejection by T lymphocytes recognizing a single amino acid difference in HLA-B44. *N Engl J Med* 1990: **323**: 1818–22.
47. Keever CA, Leong N, Cunningham I et al. HLA-B44-directed cytotoxic T cells associated with acute graft-versus-host disease following unrelated bone marrow transplantation. *Bone Marrow Transplant* 1994: **14**: 137–45.
48. Ahmad T, Neville M, Marshall SE et al. Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 2003: **12**: 647–56.
49. Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet* 2000: **67**: 947–59.
50. Single RM, Meyer D, Hollenbeck J et al. Haplotype frequency estimation in patient populations: the effect of departures from Hardy-Weinberg proportions and collapsing over a locus in the HLA region. *Genet Epidemiol* 2002: **22**: 186–95.