Histocompatibility

Increased MHC Matching by *C4* Gene Compatibility in Unrelated Donor Hematopoietic Stem Cell Transplantation

Jonna Clancy[1], Jarmo Ritari[1], Muriel Lobier[1], Riitta Niittyvuopio[2], Urpu Salmenniemi[3], Mervi Putkonen[3], Maija Itälä-Remes[3], Jukka Partanen[1], Satu Koskela[1,*]

[1] *Research and Development, Finnish Red Cross Blood Service, Helsinki, Finland*
[2] *Comprehensive Cancer Centre, Stem Cell Transplantation Unit, Helsinki University Hospital, Helsinki, Finland*
[3] *Department of Clinical Haematology and Stem Cell Transplantation Unit, Devision of Medicine, Turku University Hospital, Finland*

A B S T R A C T

HLA matching is a prerequisite for successful allogeneic hematopoietic stem cell transplantation (HSCT) because it lowers the occurrence and severity of graft-versus-host disease (GVHD). However, matching a few alleles of the classic HLA genes only may not ensure matching of the entire MHC region. HLA haplotype matching has been reported to be beneficial in HSCT because of the variation relevant to GVHD risk in the non-HLA region. Because polymorphism in the MHC is highly population specific, we hypothesized that donors from the Finnish registry are more likely to be matched at a higher level for the Finnish patients than donors from other registries. In the present study we determined 25 single nucleotide polymorphisms (SNPs) of the complement component 4 (*C4*) gene in the γ-block segment of MHC from 115 Finnish HSCT patients and their Finnish (n = 201) and non-Finnish (n = 280) donor candidates. Full matching of HLA alleles and *C4* SNPs, independently or additively, occurred more likely in the Finnish−Finnish group as compared with the Finnish−non-Finnish group ($P < .003$). This was most striking in cases with HLA haplotypes typical of the Finnish population. Patients with ancestral HLA haplotypes (AH) were more likely to find a full HLA and *C4* matched donor, regardless of donor origin, as compared with patients without AH ($P < .0001$). Despite the clear differences at the population level, we could not find a statistical association between *C4* matching and clinical outcome. The results suggest that screening *C4* SNPs can be advantageous when an extended MHC matching or HLA haplotype matching in HSCT is required. This study also supports the need for small population-specific stem cell registries.

© 2018 American Society for Blood and Marrow Transplantation.

INTRODUCTION

Matching the classic HLA at *HLA-A, -C, -B, -DRB1,* and *-DQB1* is a prerequisite for a successful unrelated donor (URD) hematopoietic stem cell transplantation (HSCT) to evade graft-versus-host disease (GVHD), a life-threatening condition. URDs are found from volunteer stem cell registries. When adequate numbers of donors cannot be found from a national stem cell registry, donors are also searched from international registries. Depending on a patients' HLA type, usually 4 to 8 donors are requested for confirmatory HLA typing. Because HLA haplotype information is usually not available, donor selection is mainly based on allele matching at the 5 classic HLA genes together with donor age, sex, and cytomegalovirus status.

The MHC encompasses 4 Mbp of DNA sequence at 6p21.3 and is divided in 3 classes based on roles of the genes in the immune system. MHC classes I and II contain the genes of the HLA molecules that represent peptides to T cells. MHC class III, located between MHC classes I and II, includes, for example, the complement component *C4* genes. The strikingly strong linkage disequilibrium (LD) in MHC [1-5] is believed to control the diversity of haplotypes to keep functionally coordinated sets of alleles together [6]. Despite the strong LD, a few recombination hot spots are located within the MHC region [4,7-11]. Their locations have been found to be the same across populations, although some appear to be haplotype or population specific [4,8,12]. These hot spots create segmented blocks in the MHC; α-, β-, γ-, and δ-blocks containing *HLA-A, -B,* and *−C*; complement genes; and *HLA-DRB1* and *-DQB1* genes, respectively. Even though these blocks can shuffle and combine to form novel assemblies, some very fixed block combinations exist because of the strong LD [9,13-15].

The block structure and positive LD that occurs in the MHC region enable long stretches of DNA to be inherited as

ancestral haplotypes (AHs) [13,16]. Many common MHC haplotypes in whites are either AHs, some ranging from *HLA-A* up to *HLA-DQB1*, or recombinants of AHs. Thus, the complement *C4* alleles are often inherited together with the flanking *HLA-B* and *HLA-DR/DQ* alleles in the European white population [10,17-19] because of the positive LD. These conserved haplotypes of different size together with other HLA haplotypes are present at varying frequency in populations from different ethnic and/or geographic origins [20-22]. Moreover, there is variation in HLA haplotype frequencies inside distinct populations as well [23-26]. The assortment of HLA haplotypes is enormous because the frequency of the most common HLA haplotypes in a population is usually only a few percentages, and most haplotypes are found in very low frequencies [20,27].

Ethnicity of a patient affects not only the probability of finding an *HLA-A, -B, -C, -DRB1, -DQB1* allele-matched URD but also the probability of finding an HLA haplotype matched donor [27-31]. Growing evidence shows that also matching of nonclassic HLA genes, or non-HLA genes in the HLA region, is associated with better HSCT outcome [32,33], as well as matching of entire HLA haplotypes [31,34-38]. On the other hand, mismatching may be beneficial because *HLA-C* and *-DPB1* and *MICA* (*MHC class I polypeptide-related sequence A*) discrepancy has been reported to provide protection from relapsing [32,39].

The Finnish population is of mainly European genetic origin [40,41]; however, many genetic features differentiate the Finns from other Europeans [40,42] because of a relatively small founder population, historical population bottlenecks, and genetic isolation [43-47]. These events are suggested to explain the reduced HLA allele pool [26,48] and specific HLA haplotypes [49] in the Finnish population.

Based on the specific HLA haplotype spectrum in the Finnish population, we wanted to evaluate whether HSCT donor candidates to Finnish patients have different HLA haplotypes depending on the candidates' origins, regardless of the apparent classical HLA allele matching. We focused on the complement component 4 (*C4*) in the γ-block segment of the class III region, a rarely researched area of MHC in HSCT. The match status of 25 single nucleotide polymorphisms (SNPs) at the *C4* gene was evaluated in 115 Finnish HSCT patients and in all their 481 donor candidates. *C4* and HLA matching grades, independently and additively, were compared between Finnish and non-Finnish donor groups as well as the effect of mismatching on the outcome of HSCT.

METHODS

One hundred fifteen Finnish patients who had received HSCT from a registry donor between 2003 and 2016 were chosen for the study. Only patients with donor candidates from both the Finnish registry (hereafter FI donors), and other registries (hereafter non-FI donors) were selected for the study. Every donor candidate that was invited for confirmatory HLA typing (N = 481), despite known prior HLA mismatch, was included in the study. The donors represented 12 different registries; 201 were from the Finnish registry and 280 were from other registries. The study material was divided into 2 sets when appropriate: patients with putative Finnish (FI) or non-Finnish (non-FI) donors, according to donor's registry. Clinical data were available for 105 HSCT pairs. Demographic details of the study subjects and clinical outcomes of patients, including GVHD grading and relapse, are described in Table 1.

Table 1
Donor and Patient Characteristics

Donors		n	Study Subjects		Donor n	Patient n
registry	Finland	201	Age	<20	2	2
	USA	20		20-40	75	20
	Germany	242		41-60	28	57
	Poland	2		>60	0	26
	Great Britain	3	CMV	pos	60	75
	France	2		neg	45	30
	Norway	2	Gender	M	82	59
	Sweden	4		F	23	46
	Denmark	1	ABO	A	52	46
	Australia	1		B	9	12
	Canada	1		AB	7	10
	Switzerland	1		O	37	37
	non-Finnish, registry NA	1	Rh	pos	89	89
	all	**481**		neg	16	16

Patients		n	Stem Cell Source		n
diagnosis	ALL	18	Peripheral blood		91
	AML	34	Bone marrow		14
	AUL	1			
	CLL	4			
	CML	2	Clinical Outcome		n
	HL	4	no aGvHD		53
	leukemia	1	aGvHD 1-4		50
	MCL	1	aGvHD NA		2
	MDS	9	no cGvHD		52
	MDS/AML	2	cGvHD 1-2		43
	MM	15	cGvHD NA		10
	MF	6	no relapse		65
	NHL	5	relapse		37
	SAA	1	relapse NA		3
	T-ALL	1			
	T-PLL	1			
	all	**105**			

Population study; recipients n=115, donors n=481. Clinical study; n=105 recipient/donor pairs.
ALL, acute lymphocytic leukemia; AML, acute myeloid leukemia; AUL, acute undifferentiated leukemia; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; HL, Hodgkin lymphoma; MCL, mantle cell lymphoma; MDS, myelodysplastic syndrome; MM, multiple myeloma; MF, myelofibrosis; NHL, non-Hodgkin lymphoma; SAA, severe aplastic anemia; T-ALL, T-cell acute lymphoblastic leukemia; T-PLL, T-cell prolymphocytic leukemia; NA indicates not applicable.

This study was carried out in accordance with the recommendations of the Ethical Committees of Helsinki and Turku University Hospitals with written, informed consent from living patients and Finnish donors. Finnish Supervisory Authority for Welfare and Health (Valvira) granted a permit to study those from whom consent was not possible due to ask (deceased or historical subjects).

Clinical HLA Typing

Genomic DNA from the WBC fraction of whole blood or from whole blood was extracted either with QiaAmp Blood Mini kit or with QIAsymphony DSP DNA Midi Kit (Qiagen GmbH, Hilden, Germany). HLA typing was performed at the HLA laboratory of the Finnish Red Cross Blood Service, using procedures accredited by the European Federation for Immunogenetics. All patients and donor candidates of the study were typed for 1-field and 2-field resolution level by SSO (INNO-LiPA; Innogenetics Group (Fujirebio), Gent, Belgium) or rSSO-Luminex technology (Labtype, One Lambda, Inc., CA) and PCR-SSP (Micro SSP Generic HLA Class I/II DNA Typing Trays [One Lambda Inc., CA, USA] or Olerup SSP genotyping [Olerup SSP AB, Stockholm, Sweden]). Sequence-based typing for determining HLA alleles at 2-field resolution was performed with AlleleSEQR PCR/Sequencing kits (Atria Genetics, Hayward, CA, USA), using the ABI 3130xl genetic analyser (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) and the Assign 3.5+ software (Conexio Genomics Pty. Ltd., Fremantle, Australia).

C4 SNP Matching

MHC class III matching was performed with commercial Gammatype typing kit (KD-PD8.0-1(96); Conexio-Genomics, CareDx). The kit consists of a panel of 25 different primer pairs targeted at complement component C4 gene in the MHC class III region: 23 targeted primer pairs for SNPs at C4 gene and 2 primer pairs for C4A and C4B genes. The results are interpreted by the presence or absence of a particular SNP at the C4 gene without indication of zygosity. PCR conditions for each reaction were performed according to the manufacturer's instructions. The parameters for the electrophoresis were as follows: 2% agarose gel (SeaKem LE Agarose 50005, lot no. 0000576343; Lonza, Rockland, ME, USA), 150 V, and 30 to 40 minutes.

The reagent details are confidential, and therefore the particular SNPs the kit detects remain unidentified. Specific characteristics of the SNPs are not possible to provide in this study. Information on AH-defining SNPs (AH7.1, 8.1, 13.1, 18.1, 38.1, 42.1, 44.2, 44.4, 46.2, 47.1, 52.1, 54.1, 55.1, 57.1, 62.1.) was kindly provided by Dr. Bruno Vanherberghen with CareDx's approval. Because of these technical restrictions and confidentiality issues the actual haplotyping is not performed. This study provides additional information exclusively for the match grade of the C4 and HLA genes between a transplantation pair, not phasing of the genes, and therefore the results are interpreted as putative haplotype matching.

Matching Models

Five separate matching models were defined based on the number of classic HLA genes and MHC classes included (Table 2). In the 5-HLA gene matching model the match status in a putative HSCT pair was assessed according to the clinical HLA-A, -B, -C, -DRB1, and -DQB1 allele assignment (10/10 matching). The HLA-DPB1 gene was included in the 6-HLA gene matching model (12/12 matching). Differentially to the 2 HLA matching models above, SNP matching at the complement component C4 gene was applied in the C4 matching model. Finally, the 5- and 6-HLA gene matching models were combined with the C4 matching model and are hereafter referred to as 5- and 6-HLA gene haplotype models.

AH and Finnish Enriched Rare Haplotype Matching

The set of possible HLA haplotypes of the 115 patients was first constructed by reflecting a patient's HLA type and allele combination to the

Table 2
HLA and C4 Match in the Study Population

Matching Model	Match Grade	All Donors n (%)	FI Donors n (%)	Non-FI Donors n (%)
5 HLA genes				
HLA-A,-B,-C,-DRB1,-DQB1	10/10 match	399 (83.0)	179 (89.1)	220 (78.6)
n=481 subjects	any mismatch grade	82 (17.0)	22 (10.9)	60 (21.4)
	9/10 match	48 (10.0)	14 (7.0)	34 (12.1)
	8/10 match	24 (5.0)	8 (4.0)	16 (5.7)
	7/10 match	9 (1.9)	0 (0.0)	9 (3.2)
	6/10 match	1 (0.2)	0 (0.0)	1 (0.4)
6 HLA genes				
HLA-A,-B,-C,-DRB1,-DQB1,-DPB1	12/12 match	108 (23.0)	52 (11.0)	56 (11.9)
n=470 subjects	any mismatch grade	362 (77.0)	144 (30.6)	218 (46.4)
	11/12 match	241 (51.3)	113 (24.0)	128 (27.2)
	10/12 match	85 (18.1)	25 (5.3)	60 (12.8)
	9/12 match	20 (4.3)	5 (1.1)	15 (3.2)
	8/12 match	10 (1.9)	1 (0.2)	9 (1.9)
	7/12 match	4 (0.9)	0 (0.0)	4 (0.9)
	6/12 match	2 (0.4)	0 (0.0)	2 (0.4)
	no HLA-DPB1 result	11 (2.3)	5 (1.1)	6 (1.3)
C4 gene				
25 SNPs	25/25 match	263 (54.7)	155 (77.1)	108 (38.6)
n=481 subjects	any mismatch grade	218 (45.3)	46 (22.9)	172 (61.4)
	24/25 match	39 (17.9)	9 (19.6)	30 (17.4)
	23/25 match	36 (16.5)	4 (8.7)	32 (18.6)
	22/25 match	45 (20.6)	12 (26.1)	33 (19.2)
	21/25 match	26 (11.9)	6 (13.0)	20 (11.6)
	20/25 match	28 (12.8)	9 (19.6)	19 (11.0)
	19/25 match	22 (10.1)	3 (6.5)	19 (11.0)
	18/25 match	19 (8.7)	3 (6.5)	16 (9.3)
	17/25 match	3 (1.4)	0.0	3 (1.7)
putative haplotype model, 5 HLA genes				
HLA-A,-B,-C,-DRB1,-DQB1, C4	C4 match, 10/10 HLA match	250 (52.0)	150 (74.6)	100 (35.7)
n=481 subjects	any mismatch grade	231 (48.0)	51 (25.4)	180 (64.3)
	C4 match, HLA mismatch	13 (2.7)	5 (2.5)	8 (2.9)
	C4 mismatch, HLA match	149 (31.0)	29 (14.4)	120 (42.9)
	C4 mismatch, HLA mismatch	69 (14.3)	17 (8.5)	52 (18.6)
putative haplotype model, 6 HLA genes				
HLA-A,-B,-C,-DRB1,-DQB1,-DPB1, C4	C4 match, 12/12 HLA match	71 (17.8)	44 (24.6)	27 (12.3)
n=399 subjects	any mismatch grade	328 (82.2)	135 (75.4)	193 (87.7)
	C4 match, HLA mismatch	179 (44.9)	106 (59.2)	73 (33.2)
	C4 match, HLA match	37 (9.3)	8 (4.5)	29 (13.2)
	C4 mismatch, HLA mismatch	112 (28.1)	21 (11.7)	91 (41.4)

HLA-A,-B,-C,-DRB1,-DQB1 HLA-typed samples; recipients n=115, donors n=481 (FI n=201, non-FI n=280).
HLA-A,-B,-C,-DRB1,-DQB1-DPB1 HLA-typed samples; recipients n=115, donors n=470 (FI n=196, non-FI n=274).

known HLA haplotype frequencies in the Finnish population (our unpublished data). The combination of HLA haplotypes with the highest probability based on their frequency was selected as patients' putative haplotype assembly. Patients were grouped into the AH-positive set on the condition that they were positive for an AH tagging SNP and had the corresponding *HLA-B*, *-DRB1*, and *-DQB1* types. Patients negative for an AH SNP and/or did not have the corresponding HLA type were classified as AH negative, regardless of the patient−donor match status. Other classic HLA loci (*HLA-A* and *HLA-DPB1*) were ruled out of this analysis because of the known recombination sites between different genomic blocks close to these genes. Patients were also divided into Finnish enriched rare (FER)-positive or FER-negative groups comparing the putative haplotypes according to the published FER haplotypes list [49]. Because there is a known active recombination hot spot near to the *HLA-DPB1* gene, FER and AH compatibility were restricted to the 10/10 HLA matching.

Statistical Analysis

The alpha level was set at .05 for statistical tests in the population study. Chi-square tests were performed to compare HLA, *C4*, and haplotype matching between the FI and non-FI donor groups using GraphPad Prism software v.7.02 (GraphPad Software, SanDiego, CA, USA).

The effect of different mismatch/match conditions (*HLA-DPB1* and *C4*, independently and additively) on clinical outcomes of 105 patients was investigated. Patients with no acute or chronic GVHD (a/cGVHD) were compared with patients with aGVHD grades I to IV and with patients with limited or extensive forms of cGVHD. Relapsed patients were compared with nonrelapsed patients (presence/absence). For each mismatch/match condition, we computed the odds ratio (OR; Wald's unconditional maximum likelihood estimation) of observing a negative clinical outcome for patients with a mismatch compared with patients with a full match. We computed confidence intervals (CIs) using the Baptista-Pike mid-p method. We used Fisher's exact test to determine statistical significance. In addition, we carried out noninferiority testing [50] on these ORs. We defined inferiority margins on the change in proportion of negative outcome with values of $\delta = .1$ for relapse and $\delta = .25$ for aGVHD and cGVHD. From these proportion inferiority margin we computed an OR threshold for each mismatch/match condition and clinical outcome pair. The mismatch condition was considered to be noninferior to the match condition if the upper bound of the 90% CI on the OR was smaller than the OR threshold. CIs were computed using the R library *ORCI* [51].

The survival analysis was carried out by first analyzing the data using a random forest survival model to evaluate the contributions of different variables. The random forest analysis was performed using the R library *ranger* v0.10.0 [52] with default settings. The variable importance and their sampling variances were estimated through jackknife resampling. Four of the top variables (ie, donor age, patient age, *C4* match, and cGVHD) were selected for subsequent analysis with Cox and Kaplan-Meier models implemented in the R library *survival* v.2.42-3 [53]. Data were managed and plotted using the R libraries *tidyverse* v.1.2.1 [54], *data.table* v.1.10.4-3 [55], and *ggpubr* v0.1.6 [56]. The R code implementing the analyses are available in GitHub (https://github.com/FRCBS/Gammatype).

RESULTS

HLA Matching

Of the 481 patient−donor candidate pairs altogether 399 pairs (83.0%) were fully 10/10 matched for *HLA-A, -B, -C, -DRB1*, and *DQB1* genes (the 5-HLA gene matching model, Table 2). Mismatching occurred most often at *HLA-C* or *HLA-DQB1* (10% and 8.1%, respectively; data not shown). The proportion of the 10/10 HLA-matched donors was higher in the FI donor group than in the non-FI donor group (89.1% versus 78.6%; $P = .003$; OR, .45; 95% CI, .27 to .77). For seven patients (6.1%) fully 10/10 HLA-matched donors were found solely in the FI donor group. Two patients (1.7%) remained without any 10/10 HLA-matched donor candidate.

In the 6-HLA gene model, with the *HLA-DPB1* gene included, only 108 (23.0%) fully 12/12 HLA-matched pairs were found (Table 2). The proportion of HLA-matched pairs was equal in both donor groups, with 11.0% for FI donors and 11.9% for non-FI donors ($P = .4$). Thus, mismatching occurred mostly at the *HLA-DPB1* gene (73.4%) in this model.

C4 Matching

Altogether, 481 patient−donor candidate pairs were screened for the match status in the *C4* gene. Of them, 263

patient−donor candidate pairs (54.7%) were fully matched for the 25 SNPs in the *C4* gene, whereas the others were matched for 17 to 24 of the SNPs (Table 2). The proportion of full *C4* match in the FI patient−FI donor group was higher (77.1%) than in FI patient−non-FI donor group (38.6%; $P < .0001$; OR, .19; 95%, CI, .12 to .28). Distribution of the number of *C4* SNPs mismatches, however, did not differ between the 2 groups (Table 2).

Added Value of C4 Matching in HLA-Matched Patient−Donor Candidate Pairs

We analyzed whether SNPs in the *C4* gene can reveal genetic differences of the MHC in the *HLA-A, -B, -C, -DRB1*, and *-DQB1* (10/10) matched patient−donor candidate pairs (n = 399). The relative number of the fully matched pairs reduced remarkably when both HLA and *C4* matches were included; in the 5-HLA genes model, 83% of the pairs were matched, whereas the share was 52% in the 5-HLA gene haplotype model (Table 2). A full *C4* match occurred with a higher frequency in fully HLA matched FI patient−FI donor pairs (83.8%) than FI patient−non-FI donor pairs (45.5%) (Figure 1A); the difference is statistically significant ($P < .0001$; OR, .16; 95% CI, .10 to .26). Therefore, *C4* mismatching was the main differentiator between the FI and non-FI donor groups in the putative haplotype model with 5 HLA genes.

To further expand MHC matching, *HLA-DPB1* was included in the putative 6-HLA gene haplotype model, that is, 10/10 matched patient−donor candidate pairs with *HLA-DPB1* result (n = 399). Altogether, 108 pairs (27.1%) were 12/12 HLA matched, of which 71 (65.7%) were also *C4* matched (Table 2). The share of pairs with both HLA and *C4* match was higher in the FI patient−FI donor group (44/52, 84.5%) than in the non-FI donor group (27/56, 48.2%), with $P < .0001$ (OR, .17; 95% CI, .07 to .4) (Figure 1B). It is of note also that in the *DPB1* mismatch group (n = 291), most of the *C4* matched pairs were from the FI−FI group (106/179, 59.2%). The difference in distribution between the 2 donor groups was again significant (106/127 [83.5%] versus 73/164 [44.5%]; $P < .0001$; OR, .16; 95% CI, .09 to .28).

Effect of FER Haplotypes

Altogether, 38 of 115 patients (33.0%) had 1 (n = 36) or 2 (n = 2) FER haplotypes. Patients with a FER haplotype were less likely to find a 10/10 HLA matched donor ($P = .0003$; OR, 2.5; 95% CI, 1.5 to 4.1) or a 10/10 HLA and *C4* matched donor ($P = .0048$; OR, 1.76; 95% CI, 1.2 to 2.6) than patients without FER. However, patients with a FER haplotype were more likely to have 10/10 HLA and *C4* matched FI donor candidates than non-FI donor candidates ($P < .0001$, chi-square test) (Figure 2A). Thus, finding a fully 10/10 HLA and 25 *C4* SNP matched donor for a patient with a FER haplotype was highly dependent on the donor registry.

AH Matching

The most frequent Finnish HLA haplotype, AH35.2 (frequency = .08), was found homozygous in 4 patients in the study set. All 15 donor candidates (11 FI, 4 non-FI) for these 4 patients were 10/10 HLA matched and fully *C4* matched. HLA haplotype AH8.1 appeared to be homozygous in 1 patient; all 4 donor candidates (3 FI, 1 non-FI) were a full 10/10 HLA and 25 *C4* SNP match for this patient.

The AH57.1 occurred heterozygous in 3 patients and their donor candidates (n = 16). Regardless of the origin of the 16 donor candidates (2 FI, 14 non-FI), all of them were fully 10/10 HLA and 25 *C4* SNP matched.
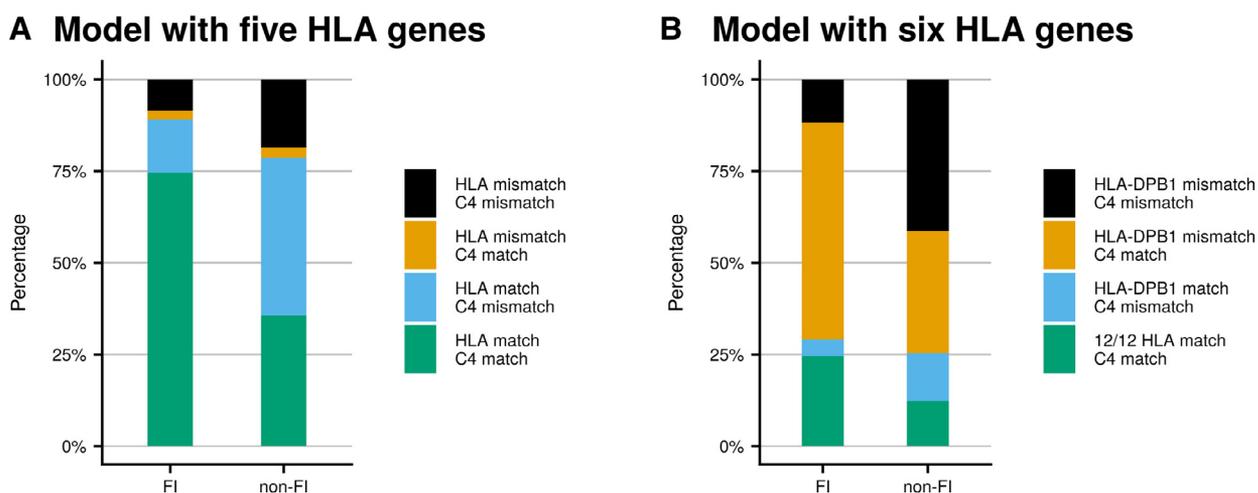
A Model with five HLA genes

B Model with six HLA genes

Figure 1. The *C4* match in the 5- and 6-HLA gene models. (A) Distribution of the *C4* match in the FI and non-FI donor groups in the 5-HLA gene matching model. (B) Distribution of the *C4* match in the FI and non-FI donor groups in the 6-HLA gene matching model. In both models the FI donors were more likely to result in a *C4* match than the non-FI donors for a Finnish patient.
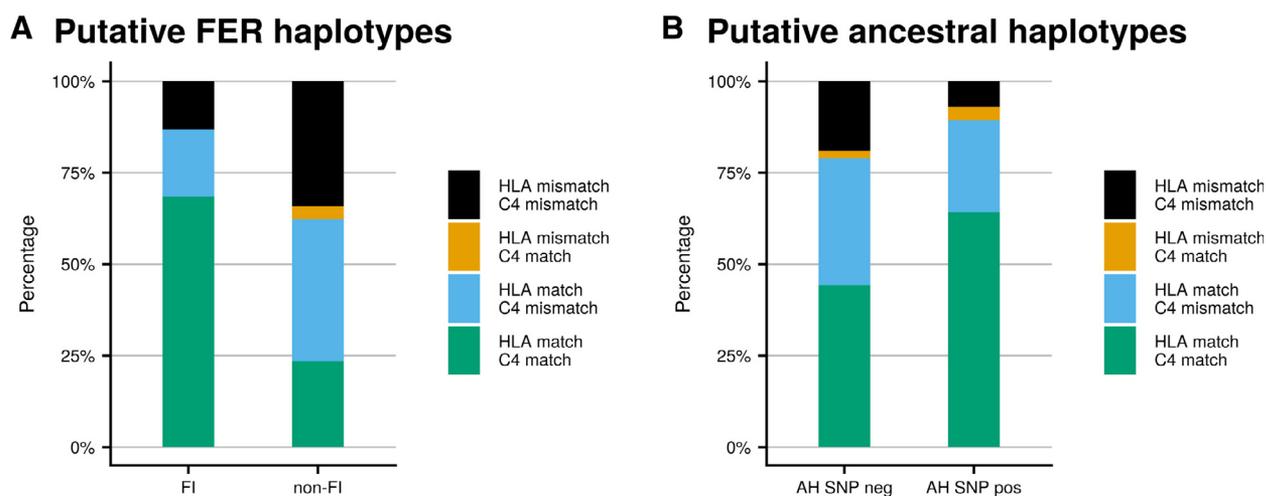
A Putative FER haplotypes

B Putative ancestral haplotypes

Figure 2. The *C4* match in the extended haplotypes. (A) Distribution of FER haplotype in FI and non-FI donor groups. Finnish patients with FER are more likely to find a 10/10 HLA matched and *C4* matched donor from the Finnish registry than from other registries. (B) Distribution of HLA match and *C4* match grades in AH-positive and -negative donor groups. Patients with AH tagging SNP are more likely to find a 10/10 HLA matched and *C4* matched donor than recipients without AH tagging SNP.

The impact of the AH in the donor search was significant (Figure 2B). Patients positive and matched for AH-associated *C4* SNPs were more likely to find 10/10 HLA matched donors than patients negative for the AH tagging SNPs ($P = .0029$; OR, 2.2; 95% Cl, 1.3 to 3.8). Patients with at least 1 AH were more likely to have a full 10/10 HLA and 25 *C4* SNP matched donor as compared with patients without any AH-specific SNPs ($P < .0001$, chi-square test).

Impact of the HLA and C4 Match on Clinical Outcome

We tested the effect HLA and *C4* matching status on clinical outcomes (cGVHD, aGVHD, and relapse) using available clinical data for 105 10/10 HLA matched HSCT. We computed the ORs (and their 95% CIs) of observing an adverse clinical outcome for patients with a mismatch compared with patients with a full match for 4 mismatch/match conditions (*HLA-DPB1* and *C4*, independently and additively). We used Fisher's exact test to test for the presence of an effect of mismatch/match status on clinical outcomes. No statistically significant differences in

clinical outcomes were found. Because the absence of a significant result does not imply the absence or presence of inferiority between mismatch/match conditions, we also ran noninferiority analyses (Supplementary Figure S1). Mismatched conditions were considered noninferior to matched conditions for relapse in the *C4* mismatch versus match condition, for cGVHD in the *HLA-DPB1* mismatch versus match condition and in the HLA-DPB1 mismatch versus match with *C4* matched condition, as well as for aGVHD in the *C4* mismatch versus match condition.

All available variables were initially screened for potential importance for survival using a random forest model. The most important variables were patient and donor age, cGVHD, diagnosis, and total matching over the 25 *C4* SNPs (GT match). These variables excluding diagnosis were selected for survival analysis using Kaplan-Meier curves and Cox regression analysis. None of the variables reached statistical significance after multiple testing adjustment, but there was a trend toward higher survival rates of about 3 years after transplantation for

patients younger than 53 years, exhibited cGVHD, or had a *C4* mismatch (Supplementary Figures S2 and S3).

DISCUSSION

It is well established that matching alleles of the classic *HLA- A, -B, -C, -DR,* and *-DQ* genes are beneficial in HSCT [57]. Several studies have suggested that increased GVHD risk after transplantation is related to *HLA-DPB1* mismatches [58,59] . Matching merely a set of classic HLA class I and II genes may not reveal possible haplotype difference in unrelated HSCT because HLA genes encompass only a small fraction of the whole MHC segment. Because haplotype data are not usually available from registry donors, the standard HLA-matched URD pairs [60,61] or even sibling pairs [38] may carry hidden mismatches in the MHC region. Matching the entire HLA haplotypes has been reported to significantly decrease the risk of GVHD and increase the overall survival in allogenic HSCT. Conversely, the incompatibility of extended MHC haplotypes significantly impairs GVHD and overall survival, emphasizing the importance of matching the entire MHC region [35,38,60-62].

In this study, based on population history and haplotype frequencies, we hypothesized that Finnish registry donors are more likely to be not only HLA matched but also HLA haplotype matched compared with non-Finnish registry donors. The small founder population, several genetic bottlenecks, and isolation by density and language have created a special genetic structure in Finns [63] and may have contributed to the MHC constitution as well. Finnish HLA haplotype frequencies are known to differ from those of neighboring populations; in fact, several common Finnish haplotypes do not exist elsewhere in Europe [26,48,49]. To identify possible haplotype matches, we used a 25 *C4* SNPs panel as the high variation both in structure and sequence of the *C4* gene [64,65], and its location at the γ-block between the β- and δ- blocks in the MHC region support its use as a haplotype determinant. In addition, the positive LD of the *C4* with its surrounding loci, *HLA-B* [17] and *HLA-DRB1* [10], further highlights its usability in haplotype matching.

According to our results, a Finnish URD is more likely to be matched with a Finnish patient than a non-Finnish donor regardless of MHC class. When each gene was individually studied, there was higher incidence of *HLA-A, -B, -C, -DRB1, -DQB1,* and *C4* matching in the FI donor group than in the non-FI group. The observed mismatches at *HLA-C* and *HLA-DQB1* are in accordance with reported haplotype differences between populations because population-specific HLA combinations are usually focused at these genes [66,67]. The idea of different HLA haplotype compositions between populations is further supported by our finding of significantly lower *C4* matching in the non-FI donor group compared with the FI group. *HLA-DPB1* was an exception because there was no difference in matching between the 2 groups. This is explained by the active recombination hotspot between *HLA-DQ* and *HLA-DP* genes [4,8,68].

The probability of finding a putative haplotype matched donor (ie, both HLA and *C4* matched) was significantly different between the 2 donor groups for the benefit of Finnish donors. We, however, underline that no real haplotyping was performed in the study because of limitations of the *C4* SNP typing method, and therefore referred haplotype models in this study are speculative. It is of note that matching was better with a Finnish donor in any match/mismatch setting as they had a higher incidence of *C4* matching despite being HLA mismatched. Our findings are congruent with the idea of the block structure of MHC where novel haplotypes are formed by shuffling the genomic blocks [14,15]. Thus, based on these results, selecting an HLA matched Finnish donor for a Finnish patient may result either in the individual MHC block match or even in the entire MHC segment match.

The putative haplotype match with an FI donor concerns especially the small group of patients who have a FER haplotype. In this group, matching reached all the way from *HLA-A* to *HLA-DQB1* and, usually, also up to the *HLA-DPB1* gene (data not shown) despite the known active recombination site just before *HLA-DPB1* [4,7,8]. Fairly limited sample size together with low effective population size [69] in the study may explain the results of relatively high number of 12/12 matches, especially in the FER group. Most isolates show substantially higher levels of LD than outbred populations [70]. The fixation of haplotypes is high in small populations because of recombination between AHs themselves [71,72]. The full matches in the FER group may also be explained simply by relatively recent introduction of these haplotypes into the Finnish population.

Even though many HLA haplotypes are population specific, some haplotypes are found to be invariable and preserved across several populations, even in distant ones. For example, the AH57.1 (HLA-A1-C6-B57-DR7-DQ3) is found in populations of European, Asian, and African origins [19,27,28,73,74]. Therefore, a fixed haplotype may give specific frames for the remaining haplotype, ensuring a haplotype match together with HLA match. This idea is supported by our findings that a proportion of HLA and *C4* matched donors is higher in the group of patients with putative AH than in the group with no AH. Also, AHs 35.2 and 8.1 were highly conserved as no variation of 25 SNPs at *C4* gene or *HLA-A, -B, -C, -DRB1,* and *-DQB1* at 2-field resolution were observed in 19 homozygous donors from both donor groups. The haplotype structure was disrupted at *HLA-DPB1* gene, as expected.

The specific MHC constitution of the Finnish population would favor a Finnish donor for a Finnish patient to minimize the risk of GVHD. Also, because nonclassic HLA and haplotype matches are reported to reduce the risk of HSCT complications [35,36,39,60-62], genetic and clinical data were combined to evaluate the effects of *C4* and haplotype mismatching on GVHD, relapse, and survival. However, we did not find any statistically significant association between *C4* compatibility and HSCT outcome, which is consistent with recent reports [75,76], although controversial results have also been reported [77]. It is of note that the impact of this MHC segment cannot completely be excluded based on our study as the relatively low number (n = 105) of actual transplantation pairs available did not afford us sufficient power to detect smaller but nevertheless clinically significant differences. A larger dataset of transplantation pairs would be required to confirm the questionable role of *C4* as such in HSCT. In any case, this study suggests that the *C4* region can be used as a HLA haplotype marker, which can be beneficial for HSCT patients because HLA haplotype matching has been reported to reduce complications after HSCT [38,60,62].

Because the frequencies of HLA haplotypes can vary greatly between populations, demand for registries representing various ethnic origins of URDs exists. When matching an unrelated registry donor to a Finnish patient before HSCT, the unique HLA constitution of Finns may display challenges. Of the Finnish patients, up to 4% find a HLA-matched donor from the Finnish Stem Cell Registry only [49]. Therefore, the need for a national stem cell registry within such a distinct population is necessary [26]. In a reverse situation, non-Finnish patients with low-frequency HLA haplotypes that are enriched in

Finland can benefit from the Finnish Stem Cell Registry. The results of this study endorse the Finnish populations' well-known role as a genetic outlier among other white populations.

ACKNOWLEDGMENTS

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.bbmt.2018.12.759.

REFERENCES

1. Dawkins R, Leelayuwat C, Gaudieri S, et al. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol Rev*. 1999;167:275−304.
2. Miretti MM, Walsh EC, Ke X, et al. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet*. 2005;76:634−646.
3. de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet*. 2006;38:1166−1172.
4. Consortium IH. A haplotype map of the human genome. *Nature*. 2005;437:1299−1320.
5. Graffelman J, Jain D, Weir B. A genome-wide study of Hardy-Weinberg equilibrium with next generation sequence data. *Hum Genet*. 2017;136: 727−741.
6. Walker BA, Hunt LG, Sowa AK, et al. The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proc Natl Acad Sci U S A*. 2011;108:8396−8401.
7. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet*. 2001;29:217−222.
8. 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;71:759−776.
9. Walsh EC, Mather KA, Schaffner SF, et al. An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet*. 2003;73:580−590.
10. Wennerström A, Vlachopoulou E, Lahtela LE, et al. Diversity of extended HLA-DRB1 haplotypes in the Finnish population. *PLoS One*. 2013;8(11): e79690.
11. Lam TH, Tay MZ, Wang B, Xiao Z, Ren EC. Intrahaplotypic variants differentiate complex linkage disequilibrium within human MHC haplotypes. *Sci Rep*. 2015;5:16972.
12. Lam TH, Shen M, Chia JM, Chan SH, Ren EC. Population-specific recombination sites within the human MHC region. *Heredity (Edinb)*. 2013;111:131−138.
13. Alper CA, Awdeh Z, Yunis EJ. Conserved, extended MHC haplotypes. *Exp Clin Immunogenet*. 1992;9:58−71.
14. Yunis EJ, Larsen CE, Fernandez-Vina M, et al. Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. *Tissue Antigens*. 2003;62:1−20.
15. Traherne JA, Horton R, Roberts AN, et al. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet*. 2006;2:e9.
16. Degli-Esposti MA, Leaver AL, Christiansen FT, Witt CS, Abraham LJ, Dawkins RL. Ancestral haplotypes: conserved population MHC haplotypes. *Hum Immunol*. 1992;34:242−252.
17. Partanen J, Koskimies S. Human MHC class III genes, Bf and C4. Polymorphism, complotypes and association with MHC class I genes in the Finnish population. *Hum Hered*. 1986;36:269−275.
18. Truedsson L, Awdeh Z, Yunis EJ, Mrose S, Moore B, Alper CA. Quantitative variation of C4 variant proteins associated with many MHC haplotypes. *Immunogenetics*. 1989;30:414−421.
19. Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA, Yunis EJ. The haplotype structure of the human major histocompatibility complex. *Hum Immunol*. 67:73-84.
20. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol*. 2007;68:779−788.
21. Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Hum Immunol*. 2013;74:1313−1320.
22. Zhou XY, Zhu FM, Li JP, et al. High-resolution analyses of human leukocyte antigens allele and haplotype frequencies based on 169, 995 volunteers from the China bone marrow donor registry program. *PLoS One*. 2015;10:1−26.
23. Nowak J, Mika-Witkowska R, Polak M, et al. Allele and extended haplotype polymorphism of HLA-A, -C, -B, -DRB1 and -DQB1 loci in Polish population and genetic affinities to other populations. *Tissue Antigens*. 2008;71:193−205.
24. Schmidt AH, Solloch UV, Baier D, et al. Regional differences in HLA antigen and haplotype frequency distributions in Germany and their relevance to the optimization of hematopoietic stem cell donor recruitment. *Tissue Antigens*. 2010;76:362−379.
25. Nicoloso G, Tiercy J, Sanchez-mazas A. *The heterogeneous HLA genetic makeup of the Swiss population*. 2012;7:1−12.
26. Sirén MK, Sareneva H, Lokki ML, Koskimies S. Unique HLA antigen frequencies in the Finnish population. *Tissue Antigens*. 1996;48:703−707.
27. Gourraud PA, Khankhanian P, Cereb N, et al. HLA diversity in the 1000 genomes dataset. *PLoS One*. 2014;9(7):e97282.
28. Middleton D, Menchaca L, Rood H, Komerofsky R. New allele frequency database: http://www.allelefrequencies.net. *Tissue Antigens*. 2003;61: 403−407.
29. Maiers M, Gragert L, Madbouly A, et al. Global analysis of registry HLA haplotypes from 20 Million individuals: report from the IHIW Registry Diversity Group. *Int J Immunogenet*. 2013;40:66−71.
30. Pidala J, Kim J, Schell M, et al. Race/ethnicity affects the probability of finding an HLA-A, -B, -C and -DRB1 allele-matched unrelated donor and likelihood of subsequent transplant utilization. *Bone Marrow Transplant*. 2013;48:346−350.
31. Joris MM, Lankester AC, von dem Borne PA, et al. The impact of frequent HLA haplotypes in high linkage disequilibrium on donor search and clinical outcome after unrelated haematopoietic SCT. *Bone Marrow Transplant*. 2013;48:483−490.
32. Carapito R, Jung N, Kwemou M, et al. *Matching for the nonconventional MHC-I MICA gene significantly reduces the incidence of acute and chronic GVHD, Blood*. 2016;128:1979−1987.
33. Fuerst D, Neuchel C, Niederwieser D, et al. Matching for the MICA-129 polymorphism is beneficial in unrelated hematopoietic stem cell transplantation. *Blood*. 2016;128:3169−3176.
34. Petersdorf EW, Malkki M, Gooley TA, Martin PJ, Guo Z. MHC haplotype matching for unrelated hematopoietic cell transplantation. *PLoS Med*. 2007;4:0059−0068.
35. Morishima S, Ogawa S, Matsubara A, et al. Impact of highly conserved HLA haplotype on acute graft-versus-host disease. *Blood*. 2010;115:4664−4670.
36. Park Y, Cheong JW, Park MH, Kim MS, Kim JS, Kim HS. Effect of major histocompatibility haplotype matching by C4 and MICA genotyping on acute graft versus host disease in unrelated hematopoietic stem cell transplantation. *Hum Immunol*. 2016;77:176−183.
37. Kitcharoen K, Witt CS, Romphruk AV, Christiansen FT, Leelayuwat C. MICA, MICB, and MHC beta block matching in bone marrow transplantation: relevance to transplantation outcome. *Hum Immunol*. 2006;67:238−246.
38. Koskela S, Ritari J, Hyvärinen K, et al. Hidden genomic MHC disparity between HLA-matched sibling pairs in hematopoietic stem cell transplantation. *Sci Rep*. 2018;8:5396.
39. Kawase T, Matsuo K, Kashiwase K, et al. HLA mismatch combinations associated with decreased risk of relapse: implications for the molecular mechanism. *Hematology*. 2009;113:2851−2858.
40. Salmela E, Lappalainen T, Fransson I, et al. Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. *PLoS One*. 2008;3(10):e3519.
41. Guglielmino CR, Piazza A, Menozzi P, Cavalli-Sforza LL. Uralic genes in Europe. *Am J Phys Anthropol*. 1990;83:57−68.
42. Durbin RM, Altshuler DL, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061−1073.
43. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Paabo S. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci U S A*. 1996;93:12035−12039.
44. Norio R. Finnish disease heritage II: population prehistory and genetic roots of Finns. *Hum Genet*. 2003;112:457−469.
45. Kere J. Human population genetics: lessons from Finland. *Annu Rev Genomics Hum Genet*. 2001;2:103−128.
46. Palo JU, Ulmanen I, Lukka M, Ellonen P, Sajantila A. Genetic markers and population history: Finland revisited. *Eur J Hum Genet*. 2009;17:1336−1346.
47. Kerminen S, Havulinna AS, Hellenthal G, et al. Fine-scale genetic structure in Finland. *G3:Genes, Genomes, Genetics*. 2017;5:3459−3468.

48. Haimila K, Peräsaari J, Linjama T, et al. HLA antigen, allele and haplotype frequencies and their use in virtual panel reactive antigen calculations in the Finnish population. *Tissue Antigens*. 2013;81:35–43.
49. Linjama T, Eberhard H-P, Peräsaari J, Müller C, Korhonen M. A European HLA isolate and its implications for hematopoietic stem cell transplant donor procurement. *Biol Blood Marrow Transplant*. 2018;24:587–593.
50. Tunes da Silva G, Logan BR, Klein JP. Methods for equivalence and noninferiority testing. *Biol Blood Marrow Transplant*. 2009;15(1 suppl.):120–127.
51. Sun L. ORCI: Several confidence intervals for the odds ratio, R package version 1.1., https://CRAN.R-project.org/package=ORCI, 2014, Accessed September 10, 2018.
52. Wright MN, Dankowski T, Ziegler A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Stat Med*. 2017;36:1272–1284.
53. Therneau TM. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer Science & Business Media; 2000.
54. Wickham H. Tools for working with URLs and HTTP. *R package version 1.2.1*. 2016; Accessed June18, 2018.
55. Matt D, Srinivasan A. R package version 1.10.4-3; 2017, Date accesed 15.6.2018. https://libraries.io/cran/data.table/1.10.4-3. Accessed June 15, 2018.
56. Kassambara Alboukadel. *R Graphics Essentials for Great Data Visualization: 200 Practical Examples*. 1st ed. STDHA. Montpellier, France: CreateSpace Independent Publishing Platform; 2017.
57. Lee SJ, Klein J, Haagenson M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood*. 2007;110:4576–4583.
58. Crocchiolo R, Zino E, Vago L, et al. Nonpermissive HLA-DPB1 disparity is a significant independent risk factor for mortality after unrelated hematopoietic stem cell transplantation. *Blood*. 2009;114:1437–1444.
59. Martin PJ, Levine DM, Storer BE, et al. Genome-wide minor histocompatibility matching as related to the risk of graft-versus-host disease. *Blood*. 2017;129:791–798.
60. Petersdorf EW, Malkki M, Gooley TA, Martin PJ, Guo Z. MHC haplotype matching for unrelated hematopoietic cell transplantation. *PLoS Med*. 2007;4:e8.
61. Petersdorf EW, Malkki M, Horowitz MM, Spellman SR, Haagenson MD, Wang T. Mapping MHC haplotype effects in unrelated donor hematopoietic cell transplantation. *Blood*. 2013;121:1896–1905.
62. Nowak J, Nestorowicz K, Graczyk-pol E, et al. Human immunology HLA-inferred extended haplotype disparity level is more relevant than the level of HLA mismatch alone for the patients survival and GvHD in T cell-replate hematopoietic stem cell transplantation from unrelated donor. *Hum Immunol*. 2018;79:403–412.
63. Norio R, Nevanlinna HR, Perheentupa J. Hereditary diseases in Finland; rare flora in rare soul. *Ann Clin Res*. 1973;5:109–141.
64. Martinez OP, Longman-Jacobsen N, Davies R, et al. Genetics of human complement component C4 and evolution the central MHC. *Front Biosci*. 2001;6:D904–D913.
65. Belt KT, Yu CY, Carroll MC, Porter RR. Polymorphism of human complement component C4. *Immunogenetics*. 1985;21:173–180.
66. Schipper RF, D'Amaro J, Bakker JT, Bakker J, van Rood JJ, Oudshoorn M. HLA gene haplotype frequencies in bone marrow donors worldwide registries. *Hum Immunol*. 1997;52:54–71.
67. Nunes JM, Buhler S, Roessli D, Sanchez-Mazas A. HLA-net 2013 collaboration. The HLA-net G ENE [RATE] pipeline for effective HLA data analysis and its application to 145 population samples from Europe and neighbouring areas. *Tissue Antigens*. 2014;83:307–323.
68. Sanchez-Mazas A, Djoulah S, Busson M, et al. A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families. *Eur J Hum Genet*. 2000;8:33–41.
69. Charlesworth B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 2009;10:195–205.
70. Service S, DeYoung J, Karayiorgou M, et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet*. 2006;38:556–560.
71. Martin AR, Karczewski KJ, Kerminen S, et al. Haplotype sharing provides insights into fine-scale population history and disease in Finland. *Am J Hum Genet*. 2018;102:760–775.
72. Chandler CH. Parallel genome-wide fixation of ancestral alleles in partially outcrossing experimental populations of *Caenorhabditis elegans*. *Genes Genomes Genet*. 2014;4:1657–1665.
73. Goodin DS, Khankhanian P, Gourraud P-A, Vince N. Highly conserved extended haplotypes of the major histocompatibility complex and their relationship to multiple sclerosis susceptibility. *PLoS One*. 2018;13 e0190043.
74. Dorak MT, Shao W, Machulla HKG, et al. Conserved extended haplotypes of the major histocompatibility complex: further characterization. *Genes Immun*. 2006;7:450–467.
75. Moyer AM, Hashmi SK, Kroning C, et al. Human immunology does matching for SNPs in the MHC gamma block in 10 / 10 HLA-matched unrelated donor-recipient pairs undergoing allogeneic stem cell transplant improve outcomes? 2018;79:532-536.
76. Askar M, Sayer D, Wang T, et al. Analysis of single nucleotide polymorphisms (SNP) donor/recipient mismatches in the gamma block of the major histocompatibility complex (MHC) and their association with hematopoietic cell transplantation (HCT) outcomes: a CIBMTR study. *Biol Blood Marrow Transplant*. 2018;24:S354–S355.
77. Hogan H, Dimovski K, Goodridge D, Sayer D. Matching for SNP's in the MHC gamma block reduces the risk of GvHD and increases survival rates post HSCT. *Human Immunol*. 2015;76:238.