The impact of HLA-C matching on donor identification rates in a European-Caucasian population

Hans-Peter Eberhard and Carlheinz R. Müller *

Zentrales Knochenmarkspender-Register Deutschland (ZKRD), Ulm, Germany

The degree of HLA concordance with the patient has long been known to be the major donor-related prediction factor for the success of hematopoietic stem cell transplantations and, with the progress of HLA typing technology, selection criteria became more stringent with regard to the recommended loci and resolution. A late refinement was HLA-C matching, which gained broader acceptance only after the turn of the millennium. The enormous HLA polymorphism has always necessitated registries with a large number of donors in order to be able to provide well-matched donors to a substantial fraction of patients. Using a biostatistical approach, we investigated the impact of adding HLA-C at low or high resolution as a supplementary matching criterion on some key parameters in donor provision for a European-Caucasian population. Starting point is donor selection based on allele level matching for HLA-A, -B, -DRB1, and, optionally, HLA-DQB1. Without typing for HLA-C, 68% of the donors selected based on matching for HLA-A, -B, -DRB1, and -DQB1 at high resolution will also match for HLA-C, 29% will have a single and only 3% will have two HLA-C alleles different from the patient. In order to provide the same fraction of patients with a fully matched donor, a registry would have to be about twice the size if HLA-C is considered in addition to the four other loci, with the exact factor increasing with the registry's size. If the provision of donors with up to a single allele mismatch is considered, this factor doubles due to the strong linkage between HLA-B and -C. These figures only change slightly when HLA-DQB1 is completely ignored or HLA-C matching is only considered at low resolution. Our results contribute to quantifying the medical and economic impact of the progress in donor selection algorithms.

Keywords: HLA-C, haplotype frequency estimation, donor registry, hematopoietic stem cell transplantation, donor-patient matching, donor identification rate

INTRODUCTION

Allogeneic hematopoietic stem cell transplantation (HSCT) is a well-established therapy with curative potential for malignant diseases of the blood and numerous otherwise fatal non-malignant disorders [1, 2]. The degree of HLA matching with the patient has long been known to be the major donor-related factor predicting the success of a HSCT [3–5]. With the progress of HLA typing technology during the last decades, the number of loci and the typing resolution used in donor selection have continuously increased. Typing and matching for HLA-A, -B, and -DR(B1) was a long-term practice in the 1980s and 1990s when serological and cellular testing were gradually replaced by molecular technology, first for class II and later for class I, boosting the reproducibility and resolution of typing results [6, 7]. While the attention paid to HLA-DQB1 has varied, in spite of earlier indications of its relevance [8] HLA-C was largely neglected until after the turn of the millennium [9, 10]. In earlier days, the low quality of serological HLA-C testing, the strong HLA-B, -C linkage, and the habit of experts to misuse this linkage in the interpretation of inconclusive laboratory data for HLA-C hampered establishing HLA-C as transplantation antigen. After the analyses of Flomenberg, Lee, and Woolfrey [11–13] during the last decade, HLA-C is now routinely considered in donor selection [14–16], but discussions are still ongoing as to if doing so at low resolution would be sufficient [13, 17].
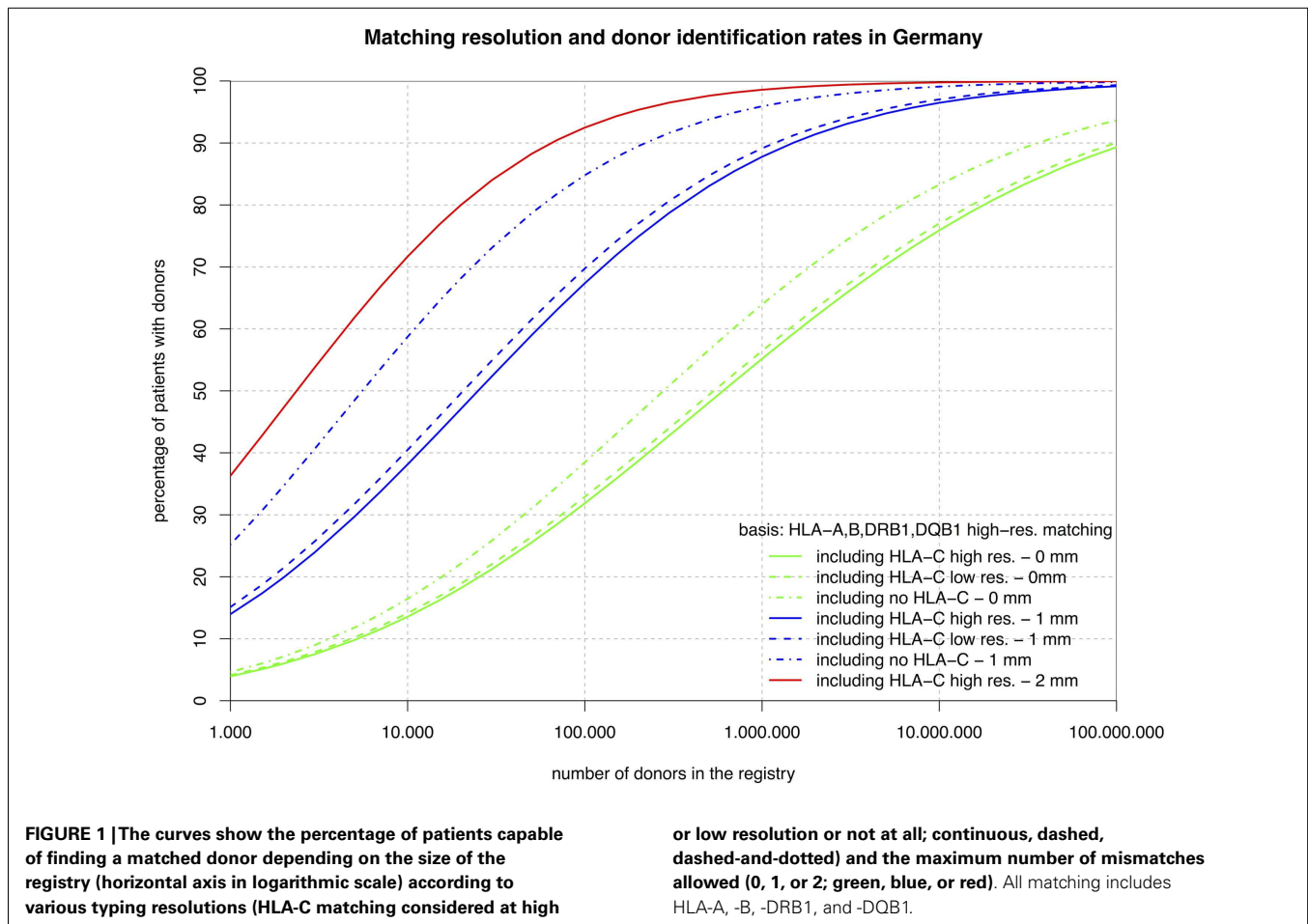
Haplotype frequency estimation (HFE) from phenotypic population data is well-established [18–20] and the challenges posed by the size of registry-derived datasets and their inherent ambiguities and heterogeneities have intensively been studied [20, 21]. In return, these haplotype frequencies (HF) can be used to study certain population properties, in particular, the dependence of the supply of donors on the size of a registry of volunteers from the population [19, 22–25].

In this study, we will investigate the impact of HLA-C matching on the donor procurement rates in a European-Caucasian population as well as the growth required for a registry to keep up its procurement rate when confronted with more stringent HLA-C matching criteria. Due to the heterogeneity of common practice with regard to HLA-DQB1 [15, 26], we will present results for matching procedures including and disregarding HLA-DQB1.

MATERIALS AND METHODS

We have estimated five-locus HLA-A, -B, -C, -DRB1, -DQB1 HF at high resolution (i.e., using the first two fields of the allele names) from a snapshot of the German donor database at the ZKRD (Zentrales Knochenmarkspender-Register Deutschland, i.e., the

FIGURE 1 | The curves show the percentage of patients capable of finding a matched donor depending on the size of the registry (horizontal axis in logarithmic scale) according to various typing resolutions (HLA-C matching considered at high or low resolution or not at all; continuous, dashed, dashed-and-dotted) and the maximum number of mismatches allowed (0, 1, or 2; green, blue, or red). All matching includes HLA-A, -B, -DRB1, and -DQB1.

German National Stem Cell Donor Registry) from May 2014, including 2,001,575 individuals with HLA designations at molecular level for HLA-A, -B, -C, -DRB1, and -DQB1. All HLA assignments used are based on the IMGT/HLA Database Release 3.16.0, 2014-04-14 (27) and the World Marrow Donor Association guidelines for the use of the HLA nomenclature (28, 29).

The HFE was performed using our well-established (19, 21) and validated (30) HFE program, which was further tuned to handle the increased ambiguity and size of this sample. This allowed for the inclusion of all donors with molecular HLA typing data at any resolution, leading to a sample size of over 2 million individuals. Allelic ambiguities were dealt with as described in (21). From this dataset, the frequencies of 89,474 five-locus haplotype at high resolution have been estimated. Since variations outside the antigen recognition site (ARS, i.e., exon 2 and 3 for HLA class I and exon 2 for HLA class II) are typically not taken into account for donor selection (31–33), we have collapsed the frequencies of ARS identical haplotypes yielding a final set of 83,446 ARS-level HF.

The probability of a random patient finding at least one suitably matched donor within a given population of donors can be calculated as

$$C\left(n\right) = \sum_{i=1}^{P} p_i \times \left(1 - \left(1 - d_i\right)^n\right)$$

where P is the number of defined phenotypes, p_i is the frequency of the phenotype in the patient population, d_i is the cumulative frequency of the suitably matched donors for this phenotype in the donor population, and n is the number of donors in the registry (24). In this calculation, the frequencies p_i and d_i are easily derived from the previously obtained HF and the result $C(n)$ indicates the extent to which the patient population is "covered" by a registry of n donors. We will use the term "coverage" to refer to $C(n)$, depending on the registry size n and on the resolution of the haplotypes of the underlying HF vector. The estimation of HF as well as the calculation of phenotype frequencies from HF assumes that the examined population is reasonably close to Hardy–Weinberg equilibrium (HWE) and we know from previous studies (19, 21) that the German donor population does not substantially deviate from the HWE assumption.

In order to analyze various matching scenarios, from the full set of 83,446 HF, we derived several marginal HF vectors by mapping HLA-C alleles to low resolution (i.e., using only the first field of the allele name) or dropping HLA-C and/or HLA-DQB1 completely from the calculations performed. Moreover, we allowed one or two mismatches in the matching criteria used to select the donors contributing to the frequencies reflected in d_i above.

All calculations were performed on an IBM x3550 server with 192 GB RAM and 24 CPU kernels running at 3.47 GHz under Ubuntu Linux 14.04. All time-critical parts of the HFE and the

coverage calculations were written in C, whereas simpler tasks, like data pre- and post-processing, were implemented in Perl. Statistical analysis was performed with supportive Perl programs, standard spreadsheet software and the open source statistical package "R" (34).

RESULTS

From the set of five-locus high-resolution HF, we have derived five related sets of marginal frequencies. The six frequency vectors are based on HLA-A, -B, -DRB1 at high resolution with C missing or added at low or high resolution and, independently, HLA-DQB1 added or omitted. For all six variants, we report the coverage based on full matches (green curves in **Figures 1** and **2**; lines numbered 1–3 in **Table 1**) and single mismatches (blue curves; lines numbered 4–6 in **Table 1**); however, we show the two mismatch coverage only for the two variants including HLA-C at high resolution (see also left part of **Table 1**).

The sets of lines with identical color in **Figures 1** and **2** illustrate the influence of adding HLA-C at low and high resolution as a matching criterion to high-resolution matching for HLA-A, -B, -DRB1, -DQB1, and HLA-A, -B, -DRB1. These graphs allow for a number of very interesting observations:

- The small difference between the dashed and the continuous lines of the same color demonstrate that the HLA-C polymorphism is already largely revealed at low resolution among individuals allele-matched for the other three or four loci.
- The large distance between the dashed-and-dotted green lines and the other two green lines shows substantial further discrimination added by HLA-C at any resolution to three- or four-locus allele matching.
- Likewise, the small difference between the corresponding green lines in the two figures reflects the strong linkage between HLA-DRB1 and -DQB1, also explaining the long ongoing debate about the relevance of HLA-DQB1 as a criterion in HSCT.
- In contrast, the substantial difference between the corresponding blue curves (one mismatch) in the two figures is quite a surprise. The reason for this is primarily, again, the strong HLA-DRB1-DQB1 linkage that shifts a substantial part of the donors mismatched for HLA-DRB1 from the blue curves of **Figure 2** (one mismatch without considering HLA-DQB1) into the higher red curve of **Figure 1** (two mismatches) and, as a consequence, brings down the blue curves.
- The red curves show that a registry of 1 million donors would provide over 98% of the patients with an 8/10 donor and over 99% with a 6/8 donor (i.e., two high-resolution mismatches, considering or ignoring HLA-DQB1). The distance from the dashed-and-dotted blue line reflects donors fully matched for HLA-C, but with two mismatches at other loci.

FIGURE 2 | The curves show the percentage of patients capable of finding a matched donor depending on the size of the registry (horizontal axis in logarithmic scale) according to various typing resolutions (HLA-C matching considered at high or low resolution or not at all; continuous, dashed, dashed-and-dotted) and the **maximum number of mismatches allowed (0, 1, or 2; green, blue, or red)**. Here, all matching includes HLA-A, -B, and -DRB1, but not HLA-DQB1.
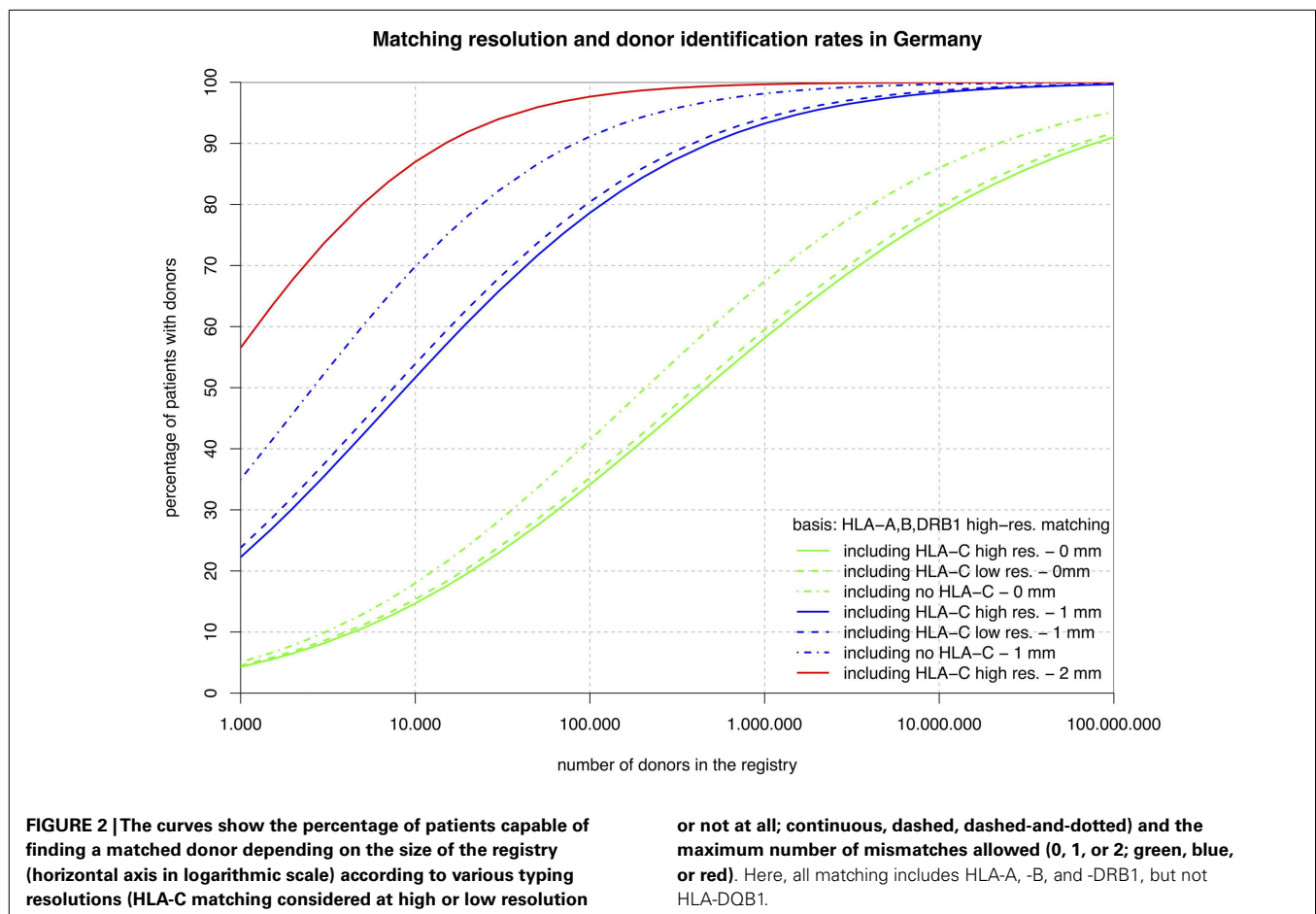
Table 1 | This table shows the attributes of the seven matching variants underlying the curves in Figures 1 and 2 and some of their key coordinates, i.e., the number of donors required to reach a coverage of 25, 50, 75, and 90% (numbers below 1,000 and over 100,000,000 have been blanked due to practical irrelevance and are also not reflected in the figures).

Curve	Matching criteria		Maximum no. of mismatches	Thousand donors required to cover patient fraction of			
	HLA-A, -B, -DRB1	HLA-C	Allowed	25%	50%	75%	90%
Figure 1							
1	Plus DQB1	High res.	0	47	602	8,879	
2	Plus DQB1	Low res.	0	42	533	7,675	98,546
3	Plus DQB1	No	0	27	277	3,251	35,763
4	Plus DQB1	High res.	1	3	25	202	1,492
5	Plus DQB1	Low res.	1	3	20	163	1,171
6	Plus DQB1	No	1	1	6	35	218
7	Plus DQB1	High res.	2		2	13	64
Figure 2							
1	No DQB1	High res.	0	38	458	6,292	77,603
2	No DQB1	Low res.	0	34	404	5,414	64,001
3	No DQB1	No	0	22	208	2,225	21,797
4	No DQB1	High res.	1	1	9	69	485
5	No DQB1	Low res.	1	1	7	56	385
6	No DQB1	No	1		3	15	82
7	No DQB1	High res.	2		1	3	14

Now, we can investigate the quantitative details of these curves and first consider the vertical distance between the green curves. Adding HLA-C matching at a registry size of 10,000 donors only reduces the coverage compared with fully matched donors by 3% points, but this gap increases to almost 10% points beyond 1 million donors.

Conversely, considering the horizontal distance between the green curves, maintaining the same coverage with the higher matching requirements demands increasing the registry size. For fully matched donors, the increment required is a factor between two and three, whereas for single-mismatched donors this factor ranges from three to around seven. Since this cannot easily be measured in the two figures due to the logarithmic scale of the x-axis, we have added the four columns on the right of **Table 1**, revealing these ratios.

As a by-product of the coverage calculations, we observed that 68% of the donors selected using high-resolution matching for HLA-A, -B, -DRB1, -DQB1 are also a full match for HLA-C, 29% will have a single mismatch and only 3% will have a double mismatch and even without including HLA-DQB1 upfront, the corresponding numbers are virtually identical (67, 30, and 3%).

DISCUSSION

The important role of HLA-C matching for a successful HSCT is well accepted today (5, 12, 13); however, the impact of a HLA-C difference for the recipient may depend on stem cell source and the type of mismatch present. The fact that over two-thirds of the donors selected on the basis of HLA-A, -B, -DRB1 high-resolution matching while omitting HLA-C are indeed HLA-C identical by chance has contributed to the success of HSCTs preceding HLA-C matching.

However, there is an important complementary aspect to this. **Figure 1** shows that, at a realistic registry size of 1 million donors, 64% of the patients have an HLA-A, -B, -DRB1, -DQB1 identical donor and 55% of the patients even have a donor that is also identical for HLA-C. Without typing for HLA-C only 44% of the patients would get a five-locus identical donor by chance. Therefore, more than half, i.e. (55–44%)/(64–44%), of the remaining mismatch situations could be avoided without increasing the registry size merely by including HLA-C as a primary selection criterion.

Maintaining the same rate of fully matched donors including HLA-C requires increasing the size of a donor registry by a factor of two to three. This may be achieved in European countries in the near future if the forthcoming revolution in HLA typing by next generation sequencing will hold its pricing promises. However, if single mismatch transplants are considered, the step from ≥7/8 to ≥9/10 matching is substantially higher since many HLA-B mismatches are typically accompanied by the corresponding HLA-C mismatch so that the size of the donor file would have to grow by a factor of three to seven to maintain the previous procurement rate.

Over 98% of the patients would have a ≥8/10 identical donor in a registry of 1 million donors so that accepting a second mismatch would almost completely resolve this problem. Unfortunately, such donors are still associated with substantially inferior outcome (5, 12, 13).

Basically, the results we have obtained underline the feasibility of including HLA-C as a primary matching criterion in donor selection for a European-Caucasian population today and, at the same time, justify and stimulate the further growth of all donor registries.

Furthermore, we were able to quantify the impact of the resolution of HLA-C typing on the identification rate of matched donors. Theoretically, in the context of matching for four other loci including HLA-B, low resolution HLA-C already reflects most of the additional polymorphism. As a consequence, it will remain difficult to decide if HLA-C matching at low resolution is sufficient or not. In practice, however, HLA-C typing should always be performed upfront for donors at the highest affordable resolution to expedite the search process and to avoid unnecessary HLA-C mismatching (17). With regard to the permissible HLA-C mismatch HLA-C*03:03 vs. 03:04 discussed in (35), a variant of the calculations reported here shows that in more than half of the cases where a low resolution HLA-C mismatch can be avoided but a high-resolution HLA-C mismatch must be accepted, this mismatch can be chosen to be HLA-C*03:03 vs. 03:04.

It is quite likely that these simulations based on a German population, with little recent admixture, will basically hold true for many similar populations of European-Caucasian background. However, it would be extremely interesting how these HLA-C-related observations translate into fundamentally different racial/ethnical groups, like those studied in (36). Unfortunately, similar investigations at the global level are still hampered by the lack of accessible multi-locus high-resolution HLA data for most populations as well as their corresponding donor availability rates. So a more realistic modeling of global donor procurement rates remains a challenge for the future.

REFERENCES

1. Thomas ED. Stem cell transplantation: past, present and future. *Stem Cells* (1994) **12**:539–44. doi:10.1002/stem.5530120602

2. Copelan EA. Hematopoietic stem-cell transplantation. *N Engl J Med* (2006) **354**:1813–26. doi:10.1056/NEJMra052638

3. Petersdorf EW, Mickelson EM, Anasetti C, Martin PJ, Woolfrey AE, Hansen JA. Effect of HLA mismatches on the outcome of hematopoietic transplants. *Curr Opin Immunol* (1999) **11**:521–6. doi:10.1016/S0952-7915(99)00016-3

4. Petersdorf EW. Optimal HLA matching in hematopoietic cell. *Curr Opin Immunol* (2008) **20**:588–93. doi:10.1016/j.coi.2008.06.014

5. Fürst D, Müller C, Vucinic V, Bunjes D, Herr W, Gramatzki M, et al. High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. *Blood* (2013) **122**:3220–9. doi:10.1182/blood-2013-02-482547

6. Ottinger H, Albert ED, Arnold R, Beelen DW, Blasczyk R, Bunjes D, et al. German consensus on immunogenetic donor search for transplantation of allogeneic bone marrow and peripheral blood stem cells. *Bone Marrow Transplant* (1997) **20**:101–5. doi:10.1038/sj.bmt.1700851

7. Ottinger HD, Müller CR, Goldmann SF, Albert ED, Arnold R, Beelen DW, et al. Second German consensus on immunogenetic donor search for allotransplantation of hematopoietic stem cells. *Ann Hematol* (2001) **80**:706–14. doi:10.1007/s002770100384-9

8. Petersdorf EW, Longton GM, Anasetti C, Mickelson EM, McKinney SK, Smith AG, et al. Association of HLA-C disparity with graft failure after marrow transplantation from unrelated donors. *Blood* (1997) **89**:1818–23.

9. Petersdorf EW, Gooley TA, Anasetti C, Martin PJ, Smith AG, Mickelson EM, et al. Optimizing outcome after unrelated marrow transplantation by comprehensive matching of HLA class I and II alleles in the donor and recipient. *Blood* (1998) **92**:3515–20.

10. Hurley CK, Baxter-Lowe LA, Logan B, Karanes C, Anasetti C, Weisdorf D, et al. National marrow donor program HLA-matching guidelines for unrelated marrow transplants. *Biol Blood Marrow Transplant* (2003) **9**:610–5. doi:10.1016/j.bbmt.2003.08.009

11. Flomenberg N, Baxter-Lowe LA, Confer DL, Fernandez-Vina MA, Filipovich AH, Horowitz MM, et al. Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. *Blood* (2004) **104**:1923–30. doi:10.1182/blood-2004-03-0803

12. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood* (2007) **110**:4576–83. doi:10.1182/blood-2007-06-097386

13. Woolfrey A, Klein JP, Haagenson M, Spellman S, Petersdorf E, Oudshoorn M, et al. HLA-C antigen mismatch is associated with worse outcome in unrelated donor peripheral blood stem cell transplantation. *Biol Blood Marrow Transplant* (2011) **17**:885–92. doi:10.1016/j.bbmt.2010.09.012

14. Bray RA, Hurley CK, Kamani NR, Woolfrey A, Müller C, Spellman S, et al. National marrow donor program HLA matching guidelines for unrelated adult donor hematopoietic cell transplants. *Biol Blood Marrow Transplant* (2008) **14**:45–53. doi:10.1016/j.bbmt.2008.06.014

15. Spellman SR, Eapen M, Logan BR, Mueller C, Rubinstein P, Setterholm MI, et al. A perspective on the selection of unrelated donors and cord blood units for transplantation. *Blood* (2012) **120**:259–65. doi:10.1182/blood-2012-03-379032

16. Tiercy J-M. HLA-C incompatibilities in allogeneic unrelated hematopoietic stem cell transplantation. *Front Immunol* (2014) **5**:216. doi:10.3389/fimmu.2014.00216

17. Dubois V, Giannoli C, Balère ML, Rey S, Raffoux C, Rigal D. Does high-resolution donor typing of HLA-C or other loci upon registration confer advantages to patients? *Hum Immunol* (2011) **72**:1033–8. doi:10.1016/j.humimm.2011.08.007

18. Yasuda N, Tsuji K. A counting method of maximum likelihood for estimating haplotype frequency in the HL-A system. *Jinrui Idengaku Zasshi* (1975) **20**:1–15.

19. Müller CR, Ehninger G, Goldmann SF. Gene and haplotype frequencies for the loci HLA-A, HLA-B, and HLA-DR based on over 13,000 German blood donors. *Hum Immunol* (2003) **64**:137–51. doi:10.1016/S0198-8859(02)00706-1

20. Kollman C, Maiers M, Gragert L, Müller CR, Setterholm M, Oudshoorn M, et al. Estimation of HLA-A, -B, -DRB1 haplotype frequencies using mixed resolution data from a National Registry with selective retyping of volunteers. *Hum Immunol* (2007) **68**:950–8. doi:10.1016/j.humimm.2007.10.009

21. Eberhard H-P, Feldmann U, Bochtler W, Baier D, Rutt C, Schmidt AH, et al. Estimating unbiased haplotype frequencies from stem cell donor samples typed at heterogeneous resolutions: a practical study based on over 1 million German donors. *Tissue Antigens* (2010) **76**:352–61. doi:10.1111/j.1399-0039.2010.01518.x

22. Beatty PG, Dahlberg S, Mickelson EM, Nisperos B, Opelz G, Martin PJ, et al. Probability of finding HLA-matched unrelated marrow donors. *Transplantation* (1988) **45**:714–8. doi:10.1097/00007890-198804000-00010

23. Kollman C, Abella E, Baitty RL, Beatty PG, Chakraborty R, Christiansen CL, et al. Assessment of optimal size and composition of the U.S. National Registry of hematopoietic stem cell donors. *Transplantation* (2004) **78**:89–95. doi:10.1097/01.TP.0000132327.40702.97

24. Sonnenberg FA, Eckman MH, Pauker SG. Bone marrow donor registries: the relation between registry size and probability of finding complete and partial matches. *Blood* (1989) **74**:2569–78.

25. Schmidt AH, Baier D, Solloch UV, Stahr A, Cereb N, Wassmuth R, et al. Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Hum Immunol* (2009) **70**:895–902. doi:10.1016/j.humimm.2009.08.006

26. Tiercy JM. Unrelated hematopoietic stem cell donor matching probability and search algorithm. *Bone Marrow Res* (2012) **2012**:695018. doi:10.1155/2012/695018

27. EMBL-EBI. *IMGT/HLA Home < IPD < EMBL-EBI* (2014). Available from: http://www.ebi.ac.uk/ipd/imgt/hla/

28. Bochtler W, Maiers M, Oudshoorn M, Marsh SGE, Raffoux C, Müller CR, et al. World Marrow Donor Association guidelines for use of HLA nomenclature

and its validation in the data exchange among hematopoietic stem cell donor registries and cord blood banks. *Bone Marrow Transplant* (2007) **39**:737–41. doi:10.1038/sj.bmt.1705672

29. Bochtler W, Maiers M, Bakker JNA, Baier DM, Hofmann JA, Pingel J, et al. An update to the HLA nomenclature guidelines of the world marrow donor association, 2012. *Bone Marrow Transplant* (2013) **48**(11):1387–8. doi:10.1038/bmt.2013.93

30. Eberhard H-P, Madbouly AS, Gourraud PA, Balère ML, Feldmann U, Gragert L, et al. Comparative validation of computer programs for haplotype frequency estimation from donor registry data. *Tissue Antigens* (2013) **82**:93–105. doi:10.1111/tan.12160

31. EMBL-EBI. Ambiguous allele combinations. *Ebiacuk* (2014). Available from: http://www.ebi.ac.uk/ipd/imgt/hla/ambig.html

32. Cano P, Klitz W, Mack SJ, Maiers M, Marsh SGE, Noreen H, et al. Common and well-documented HLA alleles: report of the Ad-Hoc committee of the American society for histocompatiblity and immunogenetics. *Hum Immunol* (2007) **68**:392–417. doi:10.1016/j.humimm.2007.01.014

33. Mack SJ, Cano P, Hollenbach JA, He J, Hurley CK, Middleton D, et al. Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens* (2013) **81**:194–203. doi:10.1111/tan.12093

34. Team RC. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available from: http://r-project.org/

35. Fernandez-Vina MA, Wang T, Lee SJ, Haagenson M, Aljurf M, Askar M, et al. Identification of a permissible HLA mismatch in hematopoietic stem cell transplantation. *Blood* (2014) **123**:1270–8. doi:10.1182/blood-2013-10-532671

36. 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**:1313–20. doi:10.1016/j.humimm.2013.06.025