Human Immunology 74 (2013) 330–340

Contents lists available at SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: How can they be used to improve unrelated donor searches?

Julia Pingel [a], Ute V. Solloch [a], Jan A. Hofmann [a], Vinzenz Lange [b], Gerhard Ehninger [c], Alexander H. Schmidt [a,b,*]

[a] DKMS German Bone Marrow Donor Center, Tübingen, Germany
[b] DKMS Life Science Lab, Dresden, Germany
[c] Internal Medicine I, University Hospital Carl Gustav Carus, Dresden, Germany

ARTICLE INFO

Article history:
Received 28 May 2012
Accepted 26 October 2012
Available online 29 November 2012

ABSTRACT

In hematopoietic stem cell transplantation, human leukocyte antigens (HLA), usually HLA loci A, B, C, DRB1 and DQB1, are required to check histocompatibility between a potential donor and the recipient suffering from a malignant or non-malignant blood disease. As databases of potential unrelated donors are very heterogeneous with respect to typing resolution and number of typed loci, donor registries make use of haplotype frequency-based algorithms to provide matching probabilities for each potentially matching recipient/donor pair. However, it is well known that HLA allele and haplotype frequencies differ significantly between populations. We estimated high-resolution HLA-A, -B, -C, -DRB1 haplotype and allele frequencies of donors within DKMS German Bone Marrow Donor Center with parentage from 17 different countries: Turkey, Poland, Italy, Russian Federation, Croatia, Greece, Austria, Kazakhstan, France, The Netherlands, Republic of China, Romania, Portugal, USA, Spain, United Kingdom and Bosnia and Herzegovina. 5-locus haplotypes including HLA-DQB1 are presented for Turkey, Poland, Italy and Russian Federation. We calculated linkage disequilibria for each sample. Genetic distances between included countries could be shown to reflect geography. We further demonstrate how genetic differences between populations are reflected in matching probabilities of recipient/donor pairs and how they influence the search for unrelated donors as well as strategic donor center typings.

1. Introduction

Allogeneic hematopoietic stem cell transplantation (HSCT) is a widely used curative therapy for the treatment of malignant diseases of the blood or other disorders of the blood and immune system [1–6]. In HSCT, compatibility of transplanted donor cells and patient's tissue depends on the human leukocyte antigen (HLA) system [7]. The HLA system is a multigenic region on chromosome 6 that shows high allelic diversity. High-resolution HLA matching is associated with improved transplantation outcome [7]. An 8/8 match between donor and recipient in HLA class I loci A, B, C and class II locus DRB1 should be found. Usually, HLA class II locus DQB1 is also considered while matching for HLA-DPB1 is not yet common practice. Unrelated stem cell donors must be typed at least at HLA-A and HLA-B and there can be all combinations of HLA typings concerning the number of typed loci and their respective typing resolutions. It is therefore challenging to find an 8/8 or 10/10 matched unrelated donor in this very heterogeneous donor pool. Typing resolution may range from serological typing to DNA-based low, intermediate or high-resolution typing. Using sequencing-based typing (SBT), high-resolution typing is often achieved; only in specific cases cis–trans typing ambiguities remain unresolved [8]. As early donor recruitment started with HLA-A and HLA-B typing only, today about 12% of almost 20 million unrelated donors worldwide are still not typed for HLA-DRB1 according to Bone Marrow Donors Worldwide (BMDW) [9].

DKMS German Bone Marrow Donor Center, for example, started SBT for HLA-A, -B, -C, -DRB1 at recruitment in 2005 and since 2010

Abbreviations: AT, Austria; BA, Bosnia and Herzegovina; CPU, central processing unit; CT, confirmatory typing; HR, Croatia; EM, expectation-maximization; FR, France; GD, genetic distance; DE, Germany; GR, Greece; HWE, Hardy–Weinberg equilibrium; HSCT, hematopoietic stem cell transplantation; HLA, human leukocyte antigen; IT, Italy; KZ, Kazakhstan; LD, linkage disequilibrium; PL, Poland; PT, Portugal; RAM, random-access memory; CN, Republic of China; RO, Romania; RU, Russian Federation; SBT, sequencing-based typing; ES, Spain; NL, The Netherlands; TR, Turkey; GB, United Kingdom; US, United States of America.

* Corresponding author. Address: DKMS German Bone Marrow Donor Center, Kressbach 1, 72072 Tübingen, Germany. Fax: +49 7071 943 1499.
E-mail address: schmidt@dkms.de (A.H. Schmidt).

0198-8859/$36.00 - see front matter © 2012 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.humimm.2012.10.029

all new donors have been typed by SBT for HLA loci A, B, C, DRB1 and DQB1.

As the search for a matching donor can be time and cost consuming, matching algorithms today start using haplotype frequencies to estimate the probability to identify a suitable 8/8 or 10/10 matched donor - thus compensating for the heterogeneity in the donor HLA database [10–12].

Thus, calculation of HLA haplotype frequencies using the expectation–maximization (EM) algorithm has become an important method in bioinformatics. We recently presented an implementation of this algorithm capable to process intermediate to high-resolution HLA data [13].

Here, we used the EM algorithm to calculate high-resolution haplotypes of donors with parentage from 17 different countries registered with DKMS German Bone Marrow Donor Center. We calculated four-locus haplotypes (HLA loci A, B, C and DRB1) and, for a subset of four countries, also five-locus haplotypes (HLA loci A, B, C, DRB1 and DQB1). As of April 2012, there were 856 different population studies including HLA allele or haplotype frequencies reported to the Allele Frequency Net database [14], however, high-resolution haplotypes for loci relevant for donor-recipient matching remain rare.

In this work, we want to demonstrate the benefits of haplotype frequencies for calculations of matching probabilities between recipients in need of a HSCT and potential donors. Taking into account the parentage of a donor in haplotype frequency-based matching calculations will further improve predictions of matching probabilities and thus facilitate the search for an unrelated stem cell donor. Along with haplotype and allele frequencies, we also present examples of matching probabilities based on our data demonstrating the differences occurring from HLA frequency differences between various populations.

2. Subjects and Methods

2.1. Donor parentage

DKMS records self-assessed parentage of each donor at recruitment, categorized by country of origin. In cases of mixed ancestry, only one country is retained which is generally the country representing the smaller part of the German population. In this study, we analyzed the HLA phenotypes of newly recruited donors by country of origin, excluding donors that were recruited in special campaigns aimed at enriching rare haplotypes in the database [15] as these donors do not represent a random selection.

For classifications of ethnic backgrounds we follow recommendations by Sanchez-Mazas et al. [16].

2.2. HLA data

DKMS donors were typed by SBT at the ASHI-accredited laboratory HistoGenetics (Ossining, NY, USA) or the EFI- and ASHI-accredited DKMS Life Science Lab (Dresden, Germany). Only very few included donors were recruited after family typing of a patient with DNA typing carried out by the corresponding transplantation center's laboratory.

HLA typing profiles at donor recruitment were analyzed and classified into three categories: (1) SBT in HLA loci A, B, C, DRB1 and DQB1 at intermediate to high-resolution (typing profile #1); (2) SBT in HLA loci A, B, C and DRB1 at intermediate to high-resolution (typing profile #2) and (3) other donors with less loci or lower resolution (typing profile #3). Donors with high-resolution typing, but yet unassigned WHO name for new alleles could not be analyzed and had to be assigned to typing profile #3.

If more than one typing result was available for a donor, as, for example, after a confirmatory typing (CT) request, the typing result with the best resolution was retained.

In total, 17 countries with at least 1000 donors typed according to profile #2 were included in the study. These countries were - in order of decreasing sample size - Turkey (TR), Poland (PL), Italy (IT), Russian Federation (RU), Croatia (HR), Greece (GR), Austria (AT), Kazakhstan (KZ), France (FR), The Netherlands (NL), Republic of China (CN), Romania (RO), Portugal (PT), United States of America (US), Spain (ES), United Kingdom (GB) and Bosnia and Herzegovina (BA). The first four countries had more than 1000 donors also typed for HLA-DQB1 (profile #1) which allowed analyses of 5-locus haplotype frequencies for these samples. Countries and sample sizes are given in Table 1.

Typing resolution was highest for class II loci yielding 94.7% high resolution, 4.7% with unresolved cis–trans ambiguities and 0.6% intermediate resolution. The corresponding values for HLA class I were 70.6%, 29.3% and 0.1%. For HLA locus A, high resolution was reached in average over all countries in 70.8% (range 65.8–73.9%), for HLA-B, HLA-C, HLA-DRB1 and HLA-DQB1 the corresponding values were 77.9% (range 71.5–83.6%), 63.1% (range 57.8–68.6%), 93.1% (range 90.7–94.7%) and 96.3% (range 95.4–97.4%), respectively.

2.3. Haplotype and allele frequencies

Haplotype frequencies were calculated using a Perl implementation of the EM algorithm validated in a previous study [13]. Calculations were carried out on a SUSE Linux Enterprise Server 10 (x86_64) with four CPU and 7.9 GB RAM.

The final sample size per country is given in Table 1. One donor with typing profile #2 and French parentage had to be excluded from high-resolution haplotype frequency analysis as the typing results were only intermediate in all loci leading to more than 1.3 million possible phenotypes and thereby exceeding the memory available for calculation. The corresponding typing result was A*02:DZMC,11:CVMB^B*35:CPVF,49:AE^C*04:DGRF,07:DZWH^DRB1*04:ASZN,12:BACJ.

For alleles identical over exon 2 and exon 3 (for class I loci HLA-A, -B, -C) and exon 2 (for class II loci HLA-DRB1, -DQB1) definitions were based on Release 3.2.0 of the ambiguous allele combinations file at the IMGT/HLA database [17]. However, we used a group representation for synonymous mutations as previously described [18]. Briefly, all alleles with identical DNA sequence over exon 2 and 3 (including null alleles) were joined with all alleles differing

Table 1
Sample sizes per country and typing profile.

Country	Country code	High resolution	
		Profile #2	Profile #1
Turkey	TR	33,083	4856
Poland	PL	6554	1761
Italy	IT	4972	1159
Russian Federation	RU	4621	1182
Croatia	HR	2057	
Greece	GR	1894	
Austria	AT	1698	
Kazakhstan	KZ	1676	
France	FR	1406	
The Netherlands	NL	1374	
Republic of China	CN	1282	
Romania	RO	1234	
Portugal	PT	1176	
United States of America	US	1111	
Spain	ES	1107	
United Kingdom	GB	1043	
Bosnia and Herzegovina	BA	1028	