Human Immunology 76 (2015) 381–384

Contents lists available at ScienceDirect

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

Rapid Communication

High-resolution HLA-A, HLA-B, and HLA-DRB1 haplotype frequencies from the French Bone Marrow Donor Registry

Pierre-Antoine Gourraud [a,b,*], Derek James Pappas [c], Amar Baouz [b], Marie-Lorraine Balère [b], Federico Garnier [b], Evelyne Marry [b]

[a] Dept. of Neurology, School of Medicine, University of California, San Francisco at San Francisco, CA, USA
[b] Registre France Greffe de Moelle, Agence de la Biomédecine, Paris, France
[c] Children's Hospital Research Institute, Oakland, CA, USA

ARTICLE INFO

Article history:
Received 6 June 2014
Accepted 21 January 2015
Available online 28 January 2015

Keywords:
Haplotype
Bone marrow donor registry
Hematopoietic stem cell donor
Peripheral blood stem cell donor cord blood bank
Genetic epidemiology

ABSTRACT

We have estimated human leukocyte antigen (HLA) haplotype frequencies using the maximum likelihood mode, which accommodates typing ambiguities. The results of the frequency distribution of the 7015 haplotypes obtained are presented here. These include a total of 114 HLA-A, 185 HLA-B, and 76 HLA-DRB1 unique alleles at each locus. Across all populations, although the most common individual HLA alleles were HLA-A*02:01 (29.0%), HLA-B*07:02 (11.4%), and HLA-DRB1*07:01 (15.9%), the most frequent haplotype was found to be HLA-A*01:01∼B*08:01∼DRB1*03:01.

1. Introduction

Estimations of human leukocyte antigen (HLA) haplotype frequencies from unrelated hematopoietic stem cell donor registries provide valuable results because they are estimated from the largest sample available. Although large-scale research sampling in the genomic era [1–3] and in large population cohorts of epidemiological interest have been collected (B58 cohort) [4,5], none of these compare the sample sizes required to find a suitable match for patients with an indication for hematopoietic stem cell transplantation (HSCT) [6]. At the national and international levels, it is estimated that data for more than 25 million hematopoietic stem cell donors worldwide have been gathered [7,8]. Although registries have a collective mission to maximize the chance of finding allogeneic sources of hematopoietic stem cells by providing access to all adult donors and cord blood units available, their data represent a significant source of information that may also help a wide variety of research projects, including association studies in many diseases, population genetics study, evolutionary studies, and models for future therapeutic approaches.

The major focus of registries is on the diversity of HLA genes, which co-dominantly express protein products that drive the recognition of both foreign and self-derived antigens by processing and presenting peptides to the cell surface for recognition by immune cells. The assorted affinities within the binding grooves of HLA molecules control the repertoire of bound epitopes and shape immune response profiles. HLA genes located in the human major histocompatibility complex (MHC; 6p21.3) are highly polymorphic, with more than 5000 alleles identified for the class I HLA alleles alone [9]. Given the HLA alleles frequencies from any one population, random mating can result in billions of possible genotype combinations. This variability is believed to be continuously renewed by mutation and recombination, and genetic drift and selective pressure also contribute to HLA diversity [10,11].

With access to high-resolution HLA typing content from thousands of hematopoietic stem cell donors, there is an opportunity to estimate the distribution of HLA haplotypes in a population. Haplotypes are the basic unit of analysis for studying human genetics given the central role of linkage disequilibrium in modeling genetic inheritance. We report the distribution of haplotype

Abbreviations: HLA, human leukocyte antigen; HSCT, hematopoietic stem cell transplantation; MHC, major histocompatibility complex; HWE, Hardy–Weinberg equilibrium; EM, expectation–maximization.

* Corresponding author at: UCSF MS Genetics, 19A Building-NS235 UCBox-3206, Dept. of Neurology, School of Medicine, University of California, San Francisco, 675 Nelson Rising Lane, Mission Bay campus, San Francisco, CA 94158, USA. Fax: +1 415 476 5229.

E-mail address: pierreantoine.gourraud@ucsf.edu (P.-A. Gourraud).

http://dx.doi.org/10.1016/j.humimm.2015.01.028
0198-8859/

frequencies for the HLA-A, HLA-B, and HLA-DRB1 loci from 42,136 unrelated hematopoietic stem cell donors from the French Bone Marrow Donors Registry.

2. Materials and methods

2.1. Samples and typing methods

Included in this study were 42,636 unrelated hematopoietic stem cell donors from the French Bone Marrow Donor Registry. These were selected because intermediate resolution typing was used at recruitment or during the framework of the campaign that improved the average resolution of HLA typing in the respective hematopoietic stem cell donors. HLA typing was performed for HLA-A, HLA-B, and HLA-DRB1 by either sequence-based typing or sequence specific oligonucleotide probes (SSOP) techniques. All were self-reported as being of French origin at recruitment; no additional information regarding ethnic origin was collected. Actual genetic ancestry was unknown because, according to the French regulations in use, the National Registry does not allow the record to indicate ethnic origin or genetic ancestry. For a detailed description of the collection centers, please see online resource. http://www.agence-biomedecine.fr/annexes/bilan2013/donnees/cellules/03-registre/synthese.htm.

2.2. Statistics analysis

The allele frequencies were calculated by a direct counting method. Several statistics were calculated to describe the allelic and haplotype frequencies within the French population. These metrics included calculations of the mean (arithmetic and geometric) and the median haplotype frequencies, the 25th and 75th percentiles of haplotype frequencies, the number of unique haplotypes, the sum of haplotype frequencies within a given range (e.g., top 10), the minimum number of haplotypes that sum to a defined threshold (e.g., sum to 10%), and the number of haplotypes greater than or equal to a defined threshold.

2.3. Data availability

The allele and haplotype frequencies in this article are publically available at Allelfrequencies.net [12].

3. Results

3.1. HLA allele frequency

HLA frequency data are presented for 42,636 French subjects. A summary of the statistics for allele frequency data of the three most commonly typed loci (HLA-A, HLA-B, and HLA-DRB1) is presented in Table 1. There were 114 HLA-A, 185 HLA-B, and 76 HLA-DRB1 unique alleles at each locus. For each locus, the most frequent alleles ($\geqslant$5%) were as follows: for the HLA-A locus, A*02:01 (29.01%), A*01:01 (13.01%), A*03:01 (12.26%), A*24:02 (9.35%), A*29:02 (5.66%), and A*11:01 (5.00%); for the HLA-B locus, B*07:02 (11.37%), B*08:01 (9.53%), B*44:03 (8.01%), B*44:02 (7.74%), B*51:01 (7.43%), B*18:01 (5.80%), and B*15:01 (5.36%); and for HLA-DRB1, DRB1*07:01 (15.86%), DRB1*03:01 (11.84%), DRB1*15:01 (11.76%), DRB1*11:01 (8.03%), DRB1*01:01 (7.87%), DRB1*13:01 (7.66%), and DRB1*04:01 (6.47). Hardy–Weinberg equilibrium (HWE) was tested at the regional level, and no significant difference from expected equilibrium proportions was detected.

Table 1
Top 25 HLA alleles by locus.

HLA-A		HLA-B		HLA-DRB1	
Allele	Freq	Allele	Freq	Allele	Freq
02:01	0.2901	07:02	0.1137	07:01	0.1586
01:01	0.1301	08:01	0.0953	03:01	0.1184
03:01	0.1226	44:03	0.0801	15:01	0.1176
24:02	0.0935	44:02	0.0774	11:01	0.0803
29:02	0.0566	51:01	0.0743	01:01	0.0787
11:01	0.0500	18:01	0.0580	13:01	0.0766
32:01	0.0339	15:01	0.0536	04:01	0.0647
26:01	0.0298	35:01	0.0460	13:02	0.0452
23:01	0.0270	40:01	0.0356	11:04	0.0337
31:01	0.0267	14:02	0.0314	04:04	0.0273
68:01	0.0238	57:01	0.0290	14:01	0.0270
30:02	0.0180	27:05	0.0287	08:01	0.0246
25:01	0.0155	49:01	0.0230	16:01	0.0223
30:01	0.0118	35:03	0.0201	01:02	0.0156
33:01	0.0112	13:02	0.0198	04:05	0.0110
02:05	0.0094	38:01	0.0194	13:03	0.0097
68:02	0.0090	55:01	0.0158	12:01	0.0085
03:01N	0.0084	40:02	0.0152	10:01	0.0084
01:02	0.0073	50:01	0.0131	04:03	0.0083
11:02	0.0050	37:01	0.0120	01:03	0.0072
66:01	0.0034	14:01	0.0114	04:02	0.0068
29:01	0.0028	39:01	0.0106	04:07	0.0067
33:03	0.0026	58:01	0.0102	11:03	0.0067
24:02L	0.0025	35:02	0.0093	01:07	0.0064
03:02	0.0021	53:01	0.0091	09:01	0.0063

3.2. HLA haplotype frequency

HLA locus haplotype frequencies were estimated using the expectation–maximization (EM) algorithm and all ambiguities were considered in the maximum likelihood model [13]. From the 42,636 subjects typed, there are 7015 unique haplotypes representing a different haplotype in approximately every 12 subjects (Table 2). Although there are 7015 unique haplotypes, 239 represent a combination of alleles with more than 5% frequency, whereas 1346 represent a combination of alleles with less than 5% frequency. The average haplotype frequency was 0.00014 ($-\log_{10} = 3.85$), whereas the median haplotype frequency was

Table 2
Summary statistics for French HLA haplotypes.

HLA loci available		ABDRB1
Sample size (N)		42,636
Sample size ($2N$)		85,272
Number of haplotypes		7015
Sample size ($2N$) per haplotype		12
Arithmetic mean frequency		1.4E−04
Geometric mean frequency		3.5E−05
Median frequency		2.0E−05
25th percentile frequency		8.6E−05
75th percentile frequency		1.4E−04
Sum (%) of haplotype frequencies within the top	10	18.0
	25	26.1
	50	33.0
	100	40.7
	250	52.8
	500	63.1
	1000	74.1
Fewest haplotypes that sum to	10%	4
	25%	23
	50%	206
	75%	1056
Number of haplotypes with frequency	$\geqslant$0.01	7
	$\geqslant$0.005	18
	$\geqslant$0.001	123
	$\geqslant$0.0005	294
	$\geqslant$0.0001	1550