

Estimation of high-resolution 11-locus haplotype and allele frequencies based on 88,698 German potential stem cell donors

Susanne Seitz¹, Alexander H. Schmidt¹, Vinzenz Lange², Jürgen Sauter¹

¹DKMS, Tübingen, Germany

²DKMS Life Science Lab, Dresden, Germany

Introduction

Haplotype frequencies are a crucial input for registries using search algorithms to identify a matching unrelated donor for hematopoietic stem cell transplantation. Recently, DKMS expanded the typing profile for newly registered donors to include HLA-DPA1, -DQA1, -DRB3, -DRB4, and -DRB5 besides HLA-A, -B, -C, -DRB1, -DQB1, -DPB1, -E and the major histocompatibility complex (MHC) class I chain-related genes A (MICA) and B (MICB).

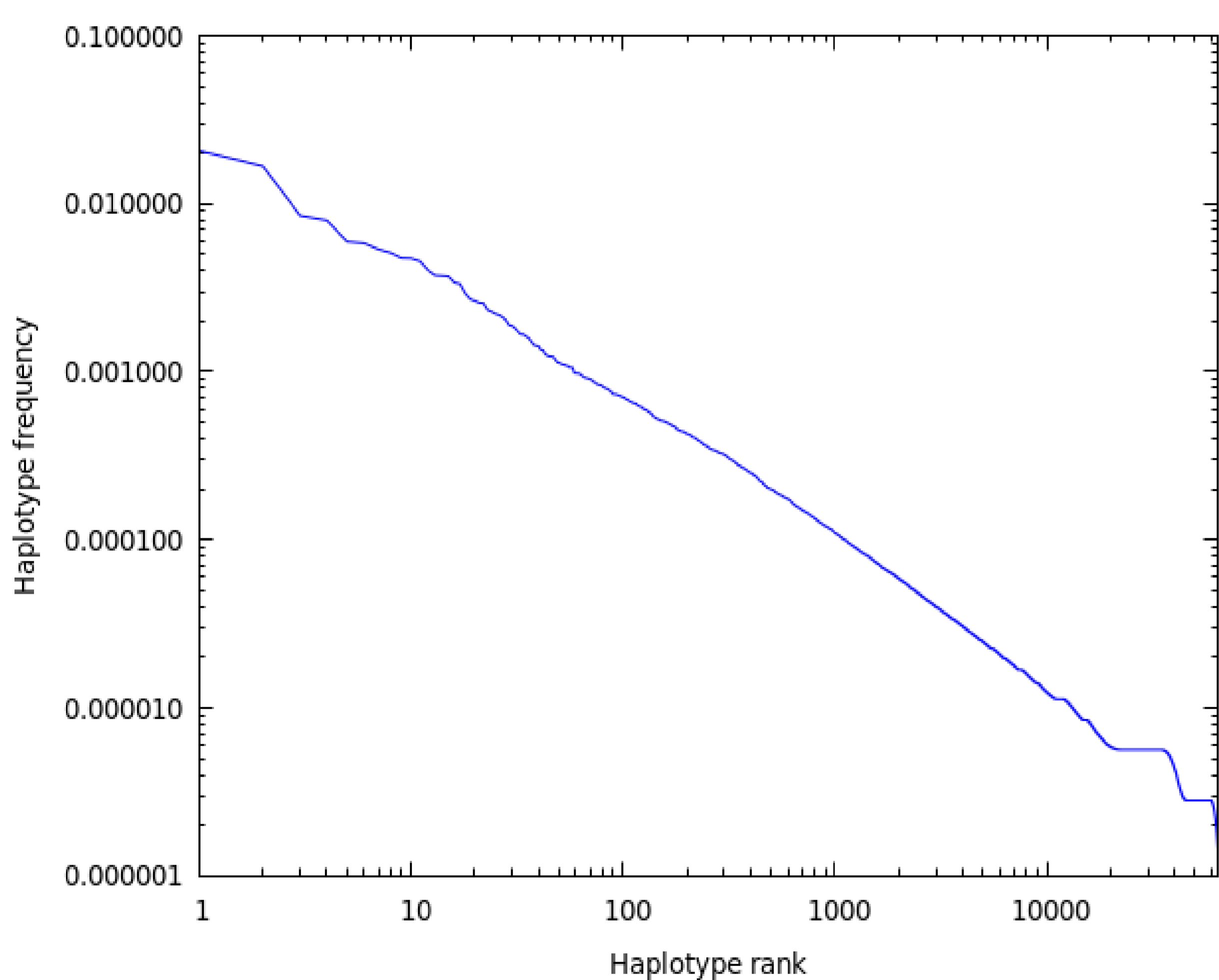
Methods

Between 10/2019 and 02/2020, n=88,698 donors with self-assigned German ethnic origin have been added to the DKMS donor file with this typing profile. All samples were typed at high-resolution at the DKMS Life-Science-Lab using our high-throughput NGS-based work flow [1,2]. For this analysis, all HLA loci are reported using the “g” nomenclature which groups all alleles that differ at most by synonymous mutations in the considered exons (2 and 3 for HLA class I and 2 for HLA class II), including null alleles [3]. For MICA and MICB, we report typing results on first-field basis and by grouping alleles that could not be distinguished by the targeted exons (2, 3 and most of exon 4)[4]. Such groups are denoted by a trailing “#” (see box below). Haplotype and allele frequencies were estimated with the EM algorithm implemented in the freely available Hapl-o-Mat software [5] (see poster P242). Haplotypes up to a cumulated frequency of 99,5% were considered. As HLA-DRB3/4/5 are not present in all genotypes these loci were disregarded in the analysis.

11-locus haplotype frequencies

Rank	Haplotype	Frequency (%)
1	A*03:01g~E*01:03g~C*07:02g~B*07:02g~MICA*008#~MICB*004#~DRB1*15:01g~DQA1*01:02g~DQB1*06:02g~DPA1*01:03g~DPB1*04:01g	2.299
2	A*01:01g~E*01:01g~C*07:01g~B*08:01g~MICA*008#~MICB*008#~DRB1*03:01g~DQA1*05:01g~DQB1*02:01g~DPA1*01:03g~DPB1*04:01g	2.066
3	A*01:01g~E*01:01g~C*07:01g~B*08:01g~MICA*008#~MICB*008#~DRB1*03:01g~DQA1*05:01g~DQB1*02:01g~DPA1*02:01g~DPB1*01:01g	1.675
4	A*02:01g~E*01:01g~C*07:02g~B*07:02g~MICA*008#~MICB*004#~DRB1*15:01g~DQA1*01:02g~DQB1*06:02g~DPA1*01:03g~DPB1*04:01g	0.849
5	A*02:01g~E*01:01g~C*05:01g~B*44:02g~MICA*008#~MICB*003#~DRB1*04:01g~DQA1*03:01g~DQB1*03:01g~DPA1*01:03g~DPB1*04:01g	0.796
6	A*03:01g~E*01:03g~C*04:01g~B*35:01g~MICA*002#~MICB*004#~DRB1*01:01g~DQA1*01:01g~DQB1*05:01g~DPA1*01:03g~DPB1*04:02g	0.595
7	A*02:01g~E*01:03g~C*03:04g~B*15:01g~MICA*010#~MICB*002#~DRB1*04:01g~DQA1*03:01g~DQB1*03:02g~DPA1*01:03g~DPB1*04:01g	0.583
8	A*03:01g~E*01:03g~C*04:01g~B*35:01g~MICA*002#~MICB*004#~DRB1*01:01g~DQA1*01:01g~DQB1*05:01g~DPA1*01:03g~DPB1*04:01g	0.536
9	A*02:01g~E*01:03g~C*03:04g~B*40:01g~MICA*008#~MICB*014#~DRB1*13:02g~DQA1*01:02g~DQB1*06:04g~DPA1*01:03g~DPB1*03:01g	0.512
10	A*02:01g~E*01:01g~C*06:02g~B*13:02g~MICA*008#~MICB*003#~DRB1*07:01g~DQA1*02:01g~DQB1*02:01g~DPA1*01:03g~DPB1*04:01g	0.476

Haplotype frequency distribution



Frequencies of the most common alleles

HLA-A	Freq. (%)	HLA-E	Freq. (%)	HLA-C	Freq. (%)	HLA-B	Freq. (%)	MICA	Freq. (%)	MICB	Freq. (%)
02:01g	28.695	01:01g	54.402	07:02g	15.343	07:02g	14.141	008#	45.286	003#	36.376
03:01g	15.883	01:03g	45.112	07:01g	14.092	08:01g	9.292	002#	11.800	004#	20.720
01:01g	14.495	01:05	0.034	04:01g	11.417	44:02g	8.329	009#	8.951	002#	13.423
24:02g	9.163	01:07	0.002	06:02g	9.636	15:01g	7.791	010#	7.661	008#	9.570
11:01g	5.246			03:04g	8.602	35:01g	6.062	004	5.796	019#	6.024
68:01g	3.690			05:01g	7.102	51:01g	5.791	007#	4.495	018	3.948
32:01g	3.647			12:03g	5.854	40:01g	5.418	018	4.038	005	3.342
26:01g	3.167			03:03g	5.845	18:01g	4.909	017#	2.915	014#	2.209
25:01g	2.556			02:02g	5.127	27:05g	3.822	012	2.300	020	2.142
31:01g	2.465			01:02g	3.549	13:02g	3.592	011	1.566	013	1.077

HLA-DRB1	Freq. (%)	HLA-DQA1	Freq. (%)	HLA-DQB1	Freq. (%)	HLA-DPA1	Freq. (%)	HLA-DPB1	Freq. (%)
15:01g	14.175	05:01g	25.173	03:01g	19.864	01:03g	83.023	04:01g	42.595
07:01g	11.868	01:02g	20.769	02:01g	18.311	02:01g	12.674	02:01g	14.261
01:01g	9.998	03:01g	15.217	06:02g	13.791	02:02g	3.212	04:02g	12.131
03:01g	9.687	01:01g	14.619	05:01g	11.911	01:04	0.497	03:01g	10.863
04:01g	8.513	02:01g	11.871	03:02g	9.691	01:05	0.094	01:01g	4.518
13:01g	7.765	01:03g	8.350	06:03g	7.982	03:01g	0.037	05:01g	1.967
11:01g	7.762	04:01g	3.159	03:03g	4.233	04:01g	0.007	13:01g	1.757
13:02g	4.264	06:01g	0.293	06:04g	3.758	01:15	0.002	17:01g	1.544
08:01g	2.968	01:10	0.051	04:02g	3.259	01:08	0.001	10:01g	1.479
11:04g	2.838	01:07g	0.037	05:03g	2.808	01:11	0.001	14:01g	1.355

MICA and MICB # groups

MICA # Group	Included MICA alleles	MICB # Group	Included MICB alleles
002#	002/020/023/052/055/089/090/091/092/093/110	002#	002/036
007#	007/026/100/111	003#	003/005:02/005:08/005:11/006/010
008#	008/027/048/087/102/103/104	004#	004/028/034
009#	009/049/109	008#	008/035/037
010#	010/065/069	014#	014/015
017#	017/095	019#	019/038

References

- [1] Schöfl, G., et al., *2.7 million samples genotyped for HLA by next generation sequencing: lessons learned*. BMC Genomics, 2017. **18**(1): p. 161.
- [2] Lange, V., et al., *Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing*. BMC Genomics, 2014. **15**: p. 63.
- [3] Schmidt, A.H., 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**(11): p. 895-902.
- [4] Klussmeier, A., Massalski, C., Putke, K., Schafer, G., Sauter, J., Schefzyk, D., . . . Lange, V. (2020). High-Throughput MICA/B Genotyping of Over Two Million Samples: Workflow and Allele Frequencies. Front Immunol, 11, 314. doi: 10.3389/fimmu.2020.00314
- [5] Schäfer, C., A.H. Schmidt, and J. Sauter, *Hapl-o-Mat: open-source software for HLA haplotype frequency estimation from ambiguous and heterogeneous data*. BMC Bioinformatics, 2017. **18**(1): p. 284.

