Immunogenetic characterization of Hungarian populations to improve the donor selection for hematopoietic stem cell transplantation

PhD thesis

Dóra Inotai

Semmelweis University
Doctoral School of Clinical Medicine

Supervisor: Attila Tordai, MD, PhD

Official reviewers: Gergely Kriván, MD, PhD
Katalin Miklós, CSc

Head of the Final Examination Committee:
Zoltán Benyó, MD, DSc
Members of the Final Examination Committee:
András Folyovich MD, PhD
Barna Vásárhelyi, MD, PhD

Budapest
2018

1 INTRODUCTION

The extent and numbers of human leukocyte antigen (HLA) mismatches between the recipient and the donor have an essential effect on the outcome of hematopoietic stem cell transplantation (HSCT). The HLA gene complex is located on the p-arm of chromosome 6. The structure of the HLA molecules are very conserved but regarding the peptid-binding domains a substantial diversity can be detected. The number of identified HLA alleles is constantly increasing due to advanced molecular biology methdos. As of December, 2017 12631 Class-I alleles and 4700 Class-II alleles have been detected. HLA-A, -B, -C, -DRB1, -DQB1 and -DPB1 are the most important loci in HSCT. Null alleles are the variants of HLA genes for which protein product is lacking or has poor cell surface expression. If a stem cell donor carries a null allele and the recipient has the variant showing normal expression, the donor derived T-cells can recognize recipient cells and induce severe acute graft versus host disease (GVHD). In the opposite situation, namely the recipient carries a null allele and there is a host versus graft mismatch, the rejection of the graft can occur. HLA-C*04:09N allele is one of the most frequent null alleles. HLA-Cw*0409N shows a single base deletion compared to wild type HLA-C*04:01 and the presence of the null allele is limited to the following haplotypes: A*02:01/A*23:01~B*44:03~DRB1*07:01~DQB1*02.

Population genetic analyses have focused on HLA genes for a few years. As a result of analyses of allele and haplotype frequencies, it has been revealed that there are some genetic differences in particular populations or a common ancestry between populations in spite of geographically far locations. In 2009, international workgroups were founded as part of HLA-NET. HLA diversity data of European populations have been published in several journals. The current study analyzing the Hungarian data fits in this process. Due to the dispersion of the Gypsies through Europe which is one of the most remarkable

movements in recent centuries, the Gypsies represent the largest minority group in Europe and in Hungary as well. Based on HLA and other studies, there was only limited genetic mixture between Gypsies and other populations therefore original characteristic features have been conserved.

There are several important clinical aspects of HLA diversity studies. In connection with several autoimmune diseases, it has been shown that particular HLA alleles have predisposing role in the pathomechanism of these diseases. Furthermore, besides the number of HLA mismatches, the preformed HLA antibodies also have a potential role in the long distance survival time of the transplanted solid organ.

During HSCT, two different recognition processes can occur due to the HLA molecules on the cell surface. Transplanted donor-derived T-cells may recognize different HLA antigens on recipient cell surface and GVHD can be induced. Certainly the immune cells of the recipient may also recognize the diverse HLA antigens of donor cells resulting host versus graft reaction (rejection). Hence the number of HLA mismatches between recipient and donor has a significant effect on HSCT outcome . Regarding the probability of GVHD and overall survival, the most suitable donor is HLA identical sibling donor, although only approximately 30% of recipients have this possibility. Therefore, national stem cell donor registries are important in cases if there is no available HLA-identical family member. Currently more than 30 million volunteer donors are listed in registries worldwide, but there are still recipients who have no suitable HLA-identical donor. In these situations, the use of alternative donor sources can be a solution such as cord blood or haploidentical family members. Despite that in recent years there has been a clear expansion in the use of haploidentical donors, registries seem to be necessary in the future. Suitable donor can most likely be found among donors who have similar ethnical background with the recipient. The reduced length of searching process can substantially contribute to improved survival.

Thus, the most primary goal of HLA diversity studies is to investigate the origin of populations and provide information of genetic mixture during migration, but it is an also important aim to facilitate the process of donor search.

2 OBJECTIVES

Our aims of HLA-C*04:09N null allele analyses were the following:

1. to establish a new method to identify HLA-C*04:09N
2. to determine the allele frequency of HLA-C*04:09N (n=7345) in the Hungarian population
3. to determine the frequency of haplotypes carrying HLA-C*04:09N in the Hungarian population

HLA frequency data is available in literature or in HLA databases for almost every European population, however DNA based HLA allele group frequency data of Hungary has not been published earlier. The primary aim of our work was to provide a new and comprehensive picture of the location of Hungary and Hungarian Gypsies on the HLA genetic map. Therefore our aims of HLA-A, -B, -DRB1 analyses were:

4. to determine HLA-A, -B, -DRB1 allele group frequencies among 2402 Hungarian volunteer hematopoietic stem cell donors
5. to define HLA-A, -B, -DRB1 allele group frequencies among 186 Hungarian Gypsy volunteer hematopoietic stem cell donors
6. to determine the following population genetics parameters in the Hungarian and Hungarian Gypsy cohorts:
    A. to assess a Hardy-Weinberg Equlibrium (HWE)
    B. to compare the detected allele group frequencies
    C. to define the cumulative allele group frequencies

D. to determine gene diversity and estimated heterozygosity

7. Our further aim was to perform a systematic comparison between Hungarian cohorts, additional European populations (the Czech Republic, Austria, Croatia, Serbia and Romania) and Gypsy-derived groups (Spanish Gypsy, North-Gujarat). Our goals to determine the genetic connections of these nine populations were the following:
   A. pairwise comparison for each locus by F-statistics based on allele group frequencies.
   B. determination and graphical presentation of the genetic distances between the nine cohorts

3 MATERIALS AND METHODS

3.1 Population samples

For HLA-C*04:09N allele frequency analyses three different groups were analyzed. The first group consists of unrelated volunteer stem cell donors (n=1953). The members of second cohort were waiting for HSCT (n=717). The third group (n=4675) was various regarding the diagnosis including recipients waiting for solid organ transplantation or HSCT.

For HLA-A, -B and –DRB1 allele frequency analyses 2402 healthy, unrelated, volunteer Hungarian stem cell donors (including 57% males and 43% females aged between 18 and 62 years) recruited between 1990 and 2012 were analyzed. This cohort represents the Hungarian average population, every member is Hungarian citizen, but it could not be excluded that few of them belong to other ethnical groups. 186 self-declared Hungarian Gypsy volunteer stem cell donors (including 60% males and 40% females aged between 18 and 62 years) recruited between 2004 and 2005 from two geographical regions, Hajdú-Bihar county (North-East of Hungary) and the area of Mohács (South of Hungary), were also analyzed. Furthermore the Czech Republic, Austria, Croatia, Serbia, Romania, Spanish Gypsy and North-Gujarat were also involved to this study. The HLA

data of non-Hungarian populations are available in literature or online HLA database.

3.2 Molecular biology methods

Low resolution genotyping was performed by PCR-SSP and PCR-SSO. In both cases, primers are responsible for amplifying exons coding antigen binding region. The identification of HLA-C*04:09N was performed by an allele specific PCR system (AS-PCR). The presence of null allele was confirmed by Sanger sequencing.

3.3 Statistical analyses

HLA-specific Gene[RATE] software was used to define HWE, HLA-A, -B, -DRB1 allele group frequencies and heterozygosity values. As a numerical indicator of gene diversity in the populations cumulative allele group frequencies were calculated and represented according to the numbers of allele groups. Pairwise comparisons of single allele group frequencies were performed by the chi-squared test using GraphPad InStat software. The degree of genetic relatedness between the nine populations was assessed by testing pairwise F_{ST}'s significance (5% level) for each locus with the Arlequin software. F-statistics is based on allele group frequencies and F_{ST} value is the population differentiation index. Reynolds genetic distances derived from pairwise F_{ST}'s. Genetic distances between the nine cohorts were graphically represented through non-metric multi-dimensional scaling analysis (NMDS) using the R statistical program. It is a data reduction method allowing us to represent genetic relationships between all objects based on pairwise genetic distances.

4 RESULTS

4.1 HLA-C*04:09N allele analyses

A new allele specific system was established by our working group to determine presence of HLA-C*04:09N allele. Since the occurrence of C*04:09N could be limited to only two haplotypes, it was sufficient that AS-PCR was performed only in cases if the patients had A*02, B*44, DRB1*07, DQB1*02 or A*23, B*44, DRB1*07, DQB1*02 allele combinations and the presence of null allele could not been excluded by serological result. In all, we found C*04:09N allele only in one case in a large dataset with 7345 persons. Based on that allele frequency in Hungarian population is 0.0068% (1/14690). The estimated haplotype frequencies for HLA-A*02-B*44-C*04-DRB1*07 and HLA-A*23-B*44-DRB1*07 haplotypes were 0.34% and 0.78% respectively.

4.2 HLA-A, -B and –DRB1 allele group analyses

4.2.1. Comparison of the Hungarian population and the Hungarian Gypsy cohort

No deviation from HWE was observed at any of the three loci in both Hungarian population (HUN) and Hungarian Gypsy group (HUN-GYP). DNA based low resolution typing data was analyzed retrospectively. At HLA-A all of the twenty different allele groups were detected in HUN and only fifteen in HUN-GYP. The two most frequent allele groups were A*02 (HUN: 29.2% and HUN-GYP: 30.4%) and A*01 (15.2% and 27.4%), the frequency of the latter was significantly higher in Gypsies ($p<0.001$). The third most common allele group was A*03 in HUN (11.8%), but not in HUN-GYP where it was much less frequent (3.5%) ($p<0.001$). In HUN-GYP, A*11 (18.8%) was the third most frequent allele group, while its frequency was significantly lower in HUN (6.0%) ($p<0.001$). Five allele groups (A*29, A*30, A*69, A*74 and A*80) found among HUN individuals were not detected among HUN-GYP.

HLA-B locus displayed the largest diversity with 28 and 26 detected allele groups in HUN and HUN-GYP, respectively. In HUN the most frequent allele groups were B*44 (12.2%), B*35 (11.3%), B*08 (10.3%), B*18 (9.6%), B*07 (8.1%), and B*51 (6.8%). By contrast, among HUN-GYP individuals, B*35 was the only allele group with a comparably high frequency (11.6%), while B*44 (5.6%) and B*08 (6.5%) frequencies were significantly lower ($p=0.0003$ and $p=0.024$). The most frequent allele groups in HUN-GYP were B*52 (14.2%), B*40 (13.4%), B*35 (11.6), B*57 (7.0%), B*08 (6.5%), and B*27 (6.2%). The two most common allele groups in HUN-GYP were significantly less frequent in HUN (2.0% and 4.4%, respectively; $p<0.001$). B*42 and B*46 was not detected neither in HUN-GYP nor in HUN. A total of 13 HLA-DRB1 allele groups were detected in HUN. The most frequent ones were DRB1*11 (15.4%), DRB1*03 (12.4%), DRB1*13 (12.2%), DRB1*07 (11.4%), DRB1*04 (11.0%), and DRB1*15 (10.6%). In HUN-GYP, the frequency of DRB1*13 (among 11 allele groups observed) was significantly lower (6.2%, $p<0.001$), while DRB1*15 (17.7%) exhibited the highest frequency followed by DRB1*03 (16.9%) and DRB1*14 (13.5%). DRB1*15 and DRB1*14 were significantly less frequent in HUN (10.6% and 5.1%, respectively, $p<0.001$). DRB1*08 and DRB1*12 allele groups could not been detected in HUN-GYP. The HLA allele group frequencies were submitted on *www.hla-net.eu* website therefore by now Hungary is on the virtual HLA map of Europe.

Analyzing the graphical representation of cumulative HLA-A, -B, -DRB1 allele frequencies for both cohorts the highest diversity and smoothest curve were observed at HLA-B. Significant difference was detected for HLA-A which exhibited steeper curve in HUN-GYP compared with that found in HUN. This is related to the fact that the sum of three most common allele group frequencies was significantly higher in HUN-GYP (76.6%) compared with HUN group (56.2%). These results were confirmed by that the estimated heterozygosity (H) was high in both populations (0.86, 0.93, and 0.89 for HLA-A, -B and -DRB1

loci in HUN, and 0.79, 0.92, and 0.87 in HUN-GYP, respectively) a significantly lower level (chi-squared test, P<0.015) was observed at HLA-A in GYP (0.79 vs 0.86 in HUN).

4.2.2. Population comparisons

F-statistics analyses were performed to define genetic relationships between two Hungarian, the other European (the Czech Republic, Austria, Croatia, Serbia, Romania) and Gypsy populations. Calculated F_{ST} and their p-values for pairwise comparisons of the populations for each locus indicated that the most genetic relationships were at HLA-A. Hungary was related to Austria and the Czech Republic (p>0.05), furthermore Austria showed relation with the Czech Republic, Croatia, Hungary and Serbia (p>0.05). For HLA-B and –DRB1 locus only two relationships were detected, namely between Austria and the Czech Republic (p>0.05) and Hungarian Gypsy and Spanish Gypsy population (p>0.05). The genetic distances between nine populations were plotted by NMDS. At the three loci, a main genetic differentiation was observed along axis 1 (MDS1) between the six non-Gypsy European populations clustered on one side and the two Gypsy plus the Indian populations on the other side. Whereas the two Gypsy populations appeared to be related to each other for HLA-B, and even more so for HLA-DRB1, they were projected far apart from the Indian population. For HLA-A Hungarian Gypsy and Spanish Gypsy cohorts located very far from each other. The HLA-A, -B and -DRB1 plots appeared to be congruent for the other European populations: Hungarians were found in an intermediate position between the Czechs and Austrians on one side and the Croats, Serbs, and Romanians, on the other side.

5 CONCLUSIONS

1. The allele frequency of HLA-C*04:09N was defined in a representative Hungarian population. Analyzing 7345 persons C*04:09N allele was

identified in only one case indicating the allele frequency is 0.0068%. This value was significantly lower compared with other published frequency data. We also estimated the haplotype frequencies for HLA-A*02-B*44-C*04-DRB1*07 (0.34%) and HLA-A*23-B*44-DRB1*07 (0.78%) haplotypes. Our results are in good agreement with data in the HLA database (*www.allelefrequencies.net*).

2. We were the first who determined allele group frequencies analyzing DNA-based low resolution HLA typing results for HLA-A, -B and –DRB1 in the Hungarian population and in the Hungarian Gypsy cohort. Comparison of the allele group frequency values showed that several allele groups were significantly more frequent in Hungarian Gypsies (A*01; A*11; B*40; B*52; DRB1*14 and DRB1*15). However, others were significantly more common in Hungarians (A*03; B*07; B*18; B*44 and DRB1*13).
3. The significant differences in HLA-A demonstrated by differences in the number of detected allele groups, the cumulative allele groups and the heterozigosity values indicated that gene diversity was lower in Hungarian Gypsies compared to the Hungarian population. These results are most probably the consequence of both a strong founder effect followed by genetic drift that occurred when Gypsies migrated out of India about 1000–1500 years. As a consequence of genetic drift is strong in small populations few allele groups have been disappeared, whereas other ones have been accumulated in the fonder population. In the Hungarian Gypsy group probably A*29 and A*30 could have been completely disappeared. Conversely the frequencies of A*02, A*01 and A*11 have been increased so much that they were three most common allele groups and the sum of the frequencies exceeded 76%. The founder effect was enhanced by isolation and endogamy in roma groups.

4. The plot of genetic distances showed that the location of Hungary was close to other European cohorts for HLA-A, -B and –DRB1. This is confirmed by clinical practice which shows the most donors for Hungarian patients come from European registries.
5. The genetic distance between the Hungarian Gypsy and Spanish Gypsy groups was small for HLA-B and –DRB1. These groups were also located close to North Gujarat cohort confirming the common origin of them. Significant difference was detected for only HLA-A, namely the three Gypsy-derived cohorts were projected far apart from each other. Consequently, HLA-A was effected by different evolution mechanisms compared to HLA-B and –DRB1.
6. Clinical consequence of our results is that for a Hungarian Gypsy recipient HLA-identical donor can be found most likely among donors with Gypsy especially Hungarian Gypsy ethnical background. To improve the chance to find a suitable donor the number of donors should be increased in the Hungarian Stem Cell Donor Registry, particularly in Gypsy minority group. Naturally Gypsy patients from all Europe could benefit from that.

6 BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

- **Publications related to the thesis:**

1. **Inotai D**, Szilvasi A, Benko S, Boros-Major A, Illes Z, Bors A, Kiss KP, Rajczy K, Gelle-Hossó A, Buhler S, Nunes JM, Sanchez-Mazas A, Tordai A. (2015) HLA genetic diversity in Hungarians and Hungarian Gypsies: Complementary differentiation patterns and demographic signals revealed by HLA-A, -B and -DRB1 in Central Europe. Tissue Antigens. 86: 115-121. **IF 2015: 2,046**
2. Bors A, **Inotai D**, Andrikovics H, Benko S, Boros-Major A, Illés Z, Szilvási A, Gelle-Hossó A, Rajczy K, Tordai A. (2015) Low occurrence of the HLA-C*04:09N allele in a large Hungarian cohort. Tissue Antigens. 86:(1) pp. 32-35. **IF 2015: 2,046**

- **Publications not related to the thesis:**

1. Balassa K, Andrikovics H, Remenyi P, Batai A, Szilvasi A, Bors A, Kiss KP, Rajczy K, **Inotai D**, Torbagyi E, Lengyel L, Barta A, Gopcsa L, Tordai A, Masszi T. (2018) Sex-specific survival difference in association with HLA-DRB1 *04 following allogeneic haematopoietic stem cell transplantation for lymphoid malignancies. Human Immunology. 79: 13-19. IF 2016: 2,311
2. Sinkovits G, Szilagyi A, Farkas P, **Inotai D**, Szilvasi A, Tordai A, Razso K, Reti M, Prohaszka Z. (2017) The role of human leukocyte antigen DRB1-DQB1 haplotypes in the susceptibility to acquired idiopathic thrombotic thrombocytopenic purpura. Human Immunology. 78:80-87. IF 2016: 2,311
3. Balassa K, Andrikovics H, Remenyi P, Batai A, Bors A, Kiss KP, Szilvasi A, Rajczy K, **Inotai D**, Gopcsa L, Lengyel L, Barta A, Reti M, Tordai A,

Masszi T. (2015) The potential role of HLA-DRB1*11 in the development and outcome of haematopoietic stem cell transplantation-associated thrombotic microangiopathy. Bone Marrow Transplantation. 50:1321-1325. IF 2015: 3,636

4. **Inotai D**, Boros-Major A, Illes Z, Szilvasi A, Nagy G, Rajczy K, Chmel R, Langer RM, Tordai A. (2012) Decrease in cold ischemic times as a result of protocol changes of urgent immunogenetic testing during cadaveric kidney transplantation in Hungary. Transplantation Proceedings. 44: 2132-2135. IF 2012: 0,952