HLA-A, -B, -C Typing by Next-Generation Sequencing in a Sample of Turkish population

Yeliz Öğret [1], F. Aytül Uyar [2], Şule Karataş [1], Fatma Savran Oğuz [1], Mehmet Tevfik Dorak [3]

[1] Dept of Medical Biology (Alp Şen Tissue Typing and Genetic Research Laboratory) &
[2] Dept of Physiology, Istanbul University, Istanbul, Turkey; [3] Liverpool Hope University, Liverpool, U.K.

SUMMARY

HLA typing was performed on 767 unrelated subjects referred to the EFI-accredited Alp Şen Tissue Typing and Genetic Research Laboratory in Istanbul University Istanbul Faculty of Medicine using Illumina MiSeq Sequencing System. The subjects were genotyped for the volunteer bone marrow donor registry. HLA-A, -B, -C typing on next-generation sequencing data was achieved using Omixon Holotype HLATM assay and Omixon HLA TwinTM software. Population genetics analyses were done on Arlequin v.3.5.1.3. Genotype frequencies at all three classical class I HLA loci were in Hardy-Weinberg equilibrium (P >= 0.25). Also, when assessed by the inbreeding coefficient (FIS), observed and expected heterozygosity did not differ at any loci (P >= 0.40). Ewens-Watterson tests of selective neutrality tests did not indicate any statistically significant selection (P = 0.99). The numbers of alleles detected in each locus were 55, 104, 51 for HLA-A, -B, -C loci, respectively. The most common three alleles (and their frequencies) at each locus were A*02:01:01 (0.201), A*24:02:01 (0.145), A*01:01:01 (0.111); B*51:01:01 (0.114), B*35:01:01 (0.081), B*18:01:01 (0.060); and C*04:01:01 (0.175), C*12:03:01 (0.114), C*07:01:01 (0.100). The most common B-C haplotypes (and their frequency / D' as LD measure) were: B*35:01:01 - C*04:01:01 (f = 0.072; D' = 0.864), B*49:01:01 - C*07:01:01 (f =0.043; D' = 0.967), B*38:01:01 - C*12:03:01 (f = 0.040; D' = 0.965). These B-C haplotypes most frequently had the following HLA-A alleles as part of three-locus haplotypes: A*11:01:01, A*23:01:01, and A*26:01:01, respectively. However, the most common three-locus haplotype was none of these, but A*03:01:01 - B*07:02:01 - C*07:02:01 (f=0.018). Of the 11 three-locus haplotypes with more than 0.01 frequency, three were the B-C haplotype B*35:01:01 - C*04:01:01 with different alleles at HLA-A: A*11:01:01 - A*24:02:01 - A*03:01:01, suggesting the presence of a recombinational hotspot between HLA-A and -C on this particular B-C haplotype. The addition of these high-resolution HLA class I types in a sample of Turkish population should fill a gap in global databases.

INTRODUCTION

HLA molecules are extremely polymorphic, with thousands of different allelic variants known in humans. Most HLA alleles show population specificity and their frequency distributions have been used in population genetics studies (1). Polymorphism is a hallmark of the HLA system, with 15 420 different HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPA1, -DPB1 alleles reported to date (release 3.26.0, October 2016) to the IPD-IMGT/HLA website.(4,5)

There is little high-resolution HLA Class I data from the Turkish population. In this preliminary study, we present our analysis of three loci (HLA-A, -B, -C) high resolution data frequencies in a sample of Turkish population.

SUBJECTS and METHODS

HLA typing was performed on 767 unrelated subjects referred to the EFI-accredited Alp Sen Tissue Typing and Genetic Research Laboratory in Istanbul University Istanbul Faculty of Medicine using Illumina MiSeq Sequencing System. The subjects were genotyped for the volunteer bone marrow donor registry. DNA was extracted from peripheral blood using Invitrogen Library Builder. Amplifications were accomplished on ABI Verity thermocycler. We sequenced at three loci (HLA-A, -B, -C) for alleles using Omixon's Holotype HLA on an Illumina Miseq using the 2x250 bp chemistry. The resulting sequencing data was analyzed and the HLA allele calls were assigned using Omixon's HLA Twin Software. Haplotype construction and linkage disequilibrium (LD) analysis was performed on Arlequin 3.5.1.3 (http://cmpg.unibe.ch/software/arlequin35).

We compared the observed allele frequencies in Turkish populations with other populations on the IDAWG Global Frequency Map Browser (http://igdawg.org/software).
We also used IMGT/HLA Allele Ethnicity Tool (http://www.ebi.ac.uk/ipd/imgt/hla/ethnicity.html) to examine ethnic-specificity of the common alleles in Turkish population.

RESULTS and DISCUSSION

The most common alleles at each locus and their frequencies were HLA-A*0201 (0.201), B*5101 (0.114), C*0401 (0.175) (Figure 1).

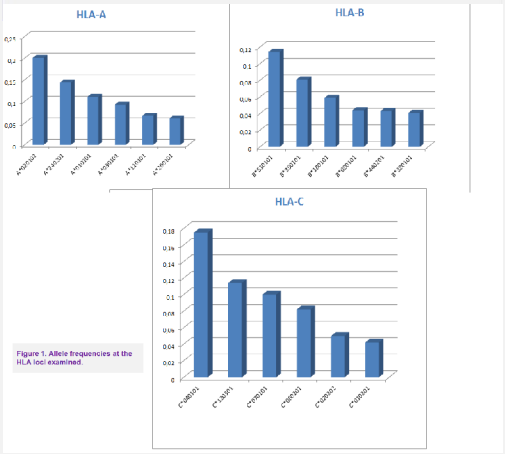


Figure 1. Allele frequencies at the HLA loci examined.

The strongest locus LD (by the delta value) was noted for B:C allele pairs, the most significant being B*35:01:01-C*04:01:01 (delta = 0.0580; *P* < 0.0000), followed by B*49:01:01-C*07:01:01 (delta = 0.0385; *P* < 0.0000) (Table 1). The statistically most significant HLA-A and B, B and C, A and C, AB and C, A and CB allele pairs, in LD (by the p value) were Table 2. Thus, the strongest LD was among the alleles of the most common European haplotype (ancestral haplotype 8.1), but it was not the most common haplotype in this sample of Turkish population.

Table 1:Strongest locus B-C LD (by the delta value)

Hyplotypes: B-C	Delta	Correlation	p
B*35:01:01-C*04:01:01	0.0580	0.8641	0.0000
B*49:01:01-C*07:01:01	0.0385	0.9673	0.0000
B*38:01:01-C*12:03:01	0.0356	0.9647	0.0000
B*35:03:01-C*04:01:01	0.0295	0.6535	0.0000
B*07:02:01-C*07:02:01	0.0319	0.9245	0.0000
B*50:01:01-C*06:02:01	0.0265	0.8532	0.0000

Table 2: A-B HYPLOTYPE LD (by the p value)

Haplotype: A-B	Delta	Correlation®	p
A*02:01:01-B*51:01:01	0.0161	0.1770	0.0000
A*03:01:01-B*07:02:01	0.0147	0.4385	0.0000
A*26:01:01-B*38:01:01	0.0156	0.4003	0.0000
A*24:02:01-B*35:01:01	0.0058	0.0833	0.0181
A*01:01:01-B*08:01:01	0.0126	0.3214	0.0000
A*02:01:01-B*18:01:01	0.0048	0.1016	0.0452
A*23:01:01-B*49:01:01	0.0147	0.4292	0.0000
A*03:01:01-B*35:01:01	0.0079	0.1082	0.0000
A*11:01:01-B*35:01:01	0.0088	0.1415	0.0000
A*24:02:01-B*35:02:01	0.0098	0.5185	0.0000
A*02:01:01-B*44:02:01	0.0043	0.1272	0.0354
A*11:01:01-B*35:03:01	0.0093	0.1826	0.0000
A*33:01:01-B*14:02:01	0.0121	0.7877	0.0000
A*03:01:01-B*18:01:01	0.0061	0.1123	0.0005
A*30:01:01-B*13:02:01	0.0112	0.5694	0.0000
A*01:01:01-B*18:01:01	0.0043	0.0825	0.0211
A*02:01:01-B*27:05:02	0.0074	0.5080	0.0000

Haplotype: A-C	Delta	Correlation®	p
A*24:02:01-C*04:01:01	0.0058	0.0823	0.0043
A*11:01:01-C*04:01:01	0.0187	0.3353	0.0000
A*02:01:01-C*07:01:01	0.0078	0.0974	0.0111
A*02:01:01-C*04:01:01	-0.008	-0.243	0.0272
A*26:01:01-C*12:03:01	0.0170	0.3108	0.0000
A*01:01:01-C*07:01:01	0.0129	0.1449	0.0000
A*01:01:01-C*07:02:01	0.0115	0.2131	0.0000
A*02:01:01-C*02:02:02	0.0100	0.2519	0.0000
A*23:01:01-C*07:01:01	0.0140	0.4340	0.0000
A*03:01:01-C*12:03:01	0.0062	0.0750	0.0083
A*01:01:01-C*06:02:01	0.0077	0.1049	0.0004
A*32:01:01-C*04:01:01	0.0073	0.1760	0.0005
A*02:01:01-C*12:03:01	-0.009	-0.404	0.0043
A*33:01:01-C*08:02:01	0.0127	0.8297	0.0000
A*30:01:01-C*06:02:01	0.0113	0.6131	0.0000
A*02:01:01-C*06:02:01	-0.005	-0.335	0.0474

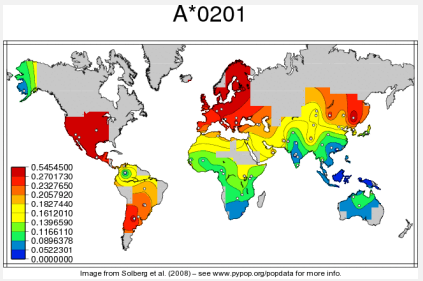
Haplotype: AC-B	Delta	Correlation®	p
A*26:01:01C*12:03:01-B*38:01:01	0.0159	0.7316	0.0000
A*24:02:01C*04:01:01-B*35:01:01	0.0139	0.4187	0.0000
A*03:01:01C*07:02:01-B*07:02:01	0.0154	0.4482	0.0000
A*23:01:01C*07:01:01-B*49:01:01	0.0153	0.7721	0.0000
A*02:01:01C*07:01:01-B*18:01:01	0.0131	0.4680	0.0000
A*11:01:01C*04:01:01-B*35:01:01	0.0105	0.3745	0.0000
A*33:01:01C*08:02:01-B*14:02:01	0.0121	0.9490	0.0000
A*02:01:01C*04:01:01-B*35:01:01	0.0095	0.3778	0.0000
A*30:01:01C*06:02:01-B*13:02:01	0.0107	0.8459	0.0000
A*11:01:01C*04:01:01-B*35:02:01	0.0087	0.3022	0.0000
A*24:02:01C*04:01:01-B*35:02:01	0.0096	0.4505	0.0000
A*01:01:01C*07:01:01-B*27:05:02	0.0088	0.4441	0.0000

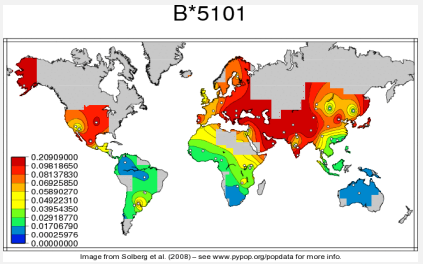
Haplotype: B-C	Delta	Correlation (r)	p
B*35:01:01-C*04:01:01	0.0580	0.8641	0.0000
B*49:01:01-C*07:01:01	0.0385	0.9673	0.0000
B*38:01:01-C*12:03:01	0.0356	0.9647	0.0000
B*35:03:01-C*04:01:01	0.0295	0.6535	0.0000
B*07:02:01-C*07:02:01	0.0319	0.9245	0.0000
B*50:01:01-C*06:02:01	0.0265	0.8532	0.0000
B*18:01:01-C*07:01:01	0.0213	0.3958	0.0000
B*51:01:01-C*15:02:01	0.0212	0.6573	0.0000
B*51:01:01-C*14:02:01	0.0209	0.8656	0.0000
B*13:02:01-C*06:02:01	0.0213	0.8909	0.0000
B*18:01:01-C*12:03:01	0.0159	0.3006	0.0000
B*08:01:01-C*07:02:01	0.0190	0.4621	0.0000
B*35:02:01-C*04:01:01	0.0182	1.0000	0.0000
B*52:01:01-C*12:02:02	0.0215	0.9430	0.0000
B*44:02:01-C*05:01:01	0.0191	0.8075	0.0000
B*08:01:01-C*07:01:01	0.0144	0.3624	0.0000
B*14:02:01-C*08:02:01	0.0172	0.9635	0.0000
B*35:08:01-C*04:01:01	0.0137	0.7503	0.0000
B*44:02:01-C*16:04:01	0.0155	0.9598	0.0000
B*39:01:01-C*12:03:01	0.0137	0.9131	0.0000
B*27:05:02-C*02:02:02	0.0140	0.8119	0.0000
B*35:03:01-C*12:03:01	0.0074	0.1534	0.0000
B*55:01:01-C*01:02:01	0.0125	0.4657	0.0000
B*41:01:01-C*17:01:01	0.0128	1.0000	0.0000
B*57:01:01-C*06:02:01	0.0117	0.8182	0.0000
B*51:01:01-C*15:13	0.0114	0.9462	0.0000
B*55:01:01-C*03:03:01	0.0115	0.4246	0.0000
B*40:01:02-C*03:04:01	0.0119	0.8210	0.0000
B*15:17:01-C*07:01:02	0.0121	0.9492	0.0000
B*58:01:01-C*03:02:02	0.0121	1.0000	0.0000
B*51:01:01-C*16:02:01	0.0086	0.6237	0.0000

Haplotype: AB-C	Delta	Correlation®	p
A*03:01:01B*07:02:01-C*07:02:01	0.0163	0.9615	0.0000
A*26:01:01B*38:01:01-C*12:03:01	0.0149	0.9581	0.0000
A*23:01:01B*49:01:01-C*07:01:01	0.0146	1.0000	0.0000
A*02:01:01B*18:01:01-C*07:01:01	0.0134	0.9536	0.0000
A*11:01:01B*35:01:01-C*04:01:01	0.0118	1.0000	0.0000
A*01:01:01B*08:01:01-C*07:01:01	0.0114	0.7776	0.0000
A*24:02:01B*35:02:01-C*04:01:01	0.0107	1.0000	0.0000
A*24:02:01B*35:01:01-C*04:01:01	0.0102	0.7978	0.0000
A*03:01:01B*35:01:01-C*04:01:01	0.0099	0.8845	0.0000
A*33:01:01B*14:02:01-C*08:02:01	0.0121	1.0000	0.0000
A*30:01:01B*13:02:01-C*06:02:01	0.0101	0.9394	0.0000

The top three most common HLA alleles at each locus were examined on IMGT/HLA Allele Ethnicity Tool, but none showed any ethnic specificity.

Implementation of high-resolution typing methods allows for the detection of a significantly wider spectrum of HLA variations. Examination of the frequencies of the most common allele in IDAWG Global Frequency Map Browser showed that the same alleles were most common in the Balkans and populations of the Near East suggesting that the contemporary Turkish population is genetically more similar to its neighbours as has been suggested before (1).
This study provided tree loci high-resolution HLA types in a sample of Turkish population which should fill the gap in global HLA allele frequency databases. Our ongoing efforts will increase the number of samples typed at high-resolution for more informative results.





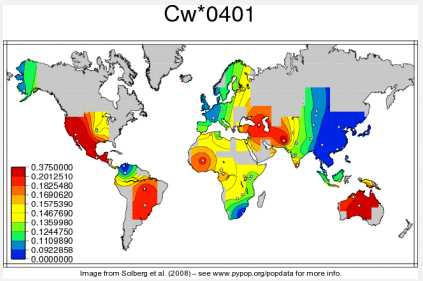


Figure 3. Global frequencies of HLA alleles most common in the Turkish population.

REFERENCES

1. Machulla HK, Batnasan D, Steinborn F, Uyar FA, Saruhan-Direskeneli G, Oguz FS, Carin MN, Dorak MT. Genetic affinities among Mongol ethnic groups and their relationship to Turks. Tissue Antigens 2003;61(4):292-9

2. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunol 2008;22;9:1

3. Dos Santos Francisco R, Buhler S, Nunes JM, Bitarello BD, França GS, Meyer D, Sanchez-Mazas A. HLA supertype variation across populations: new insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. Immunogenetics 2015;67(11-12):651-63.

4. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43:D423–D431.

5. Robinson J, Halliwell JA, Marsh SG. IMGT/HLAand the immuno polymorphism database. Methods Mol Biol. 2014;1184:109–121.