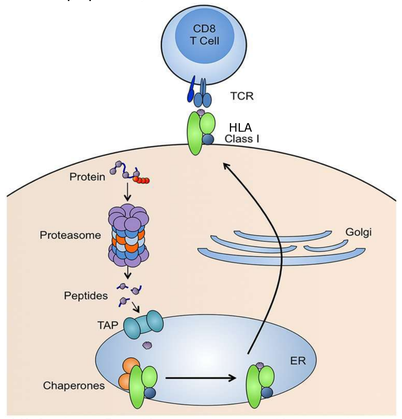
Variability of immunologically significant regions of HLA class I proteins

Arsen Tedeev[1], Vadim Karnaukhov[2], Ivan Zvyagin[3]

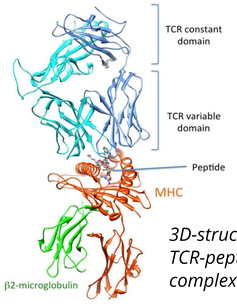
[1]*Faculty of Bioingeneering and Bioinformatics, Lomonosov Moscow State University;* [2]*Skolkovo Institute of Science and Technology;*
[3]*Laboratory of Comparative and Structural Genomics, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry*

Introduction

- Human leukocyte antigen (HLA) proteins are key components of the immune system
- HLA class I proteins bind peptides derived from degraded intracellular proteins and present them on the cell surface for inspection by T-cell receptors (TCR) of T-cells
- TCR interacts both with HLA and peptide
- TCR recognition of pathogenic peptide presented by HLA triggers adaptive immune response
- HLA class I genes are HLA-A, HLA-B and HLA-C
- HLA genes are the most polymorphic in the genome (>10000 variants, presenting different sets of peptides)



HLA class I antigen presentation pathway



3D-structure of TCR-peptide-MHC complex

Aims

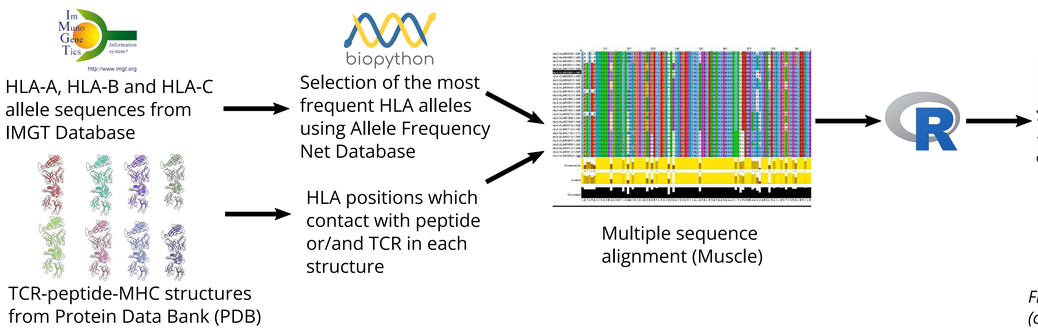
1. To analyse the variability of TCR- and peptide-contacting regions of HLA.
2. To investigate variability within HLA-A, -B and -C groups of alleles.
3. To search for evolutionary conservative sequences of peptide- or/and TCR-contacting positions.

Conclusions

1.Peptide-contacting positions are more polymorphic than TCR-contacting ones. Non-contacting positions are conservative.
2. HLA-A alleles differ mostly in TCR-contacting positions, HLA-B - in peptide-contacting ones.
3. We found a conservative sequence of TCR-contacting residues in a set of HLA-B alleles which may be evolved convergently.

Results

1. Identification of peptide- and TCR-contacting regions of HLA



Regions with a higher degree of variability coincide with TCR- and peptide-contacting regions.

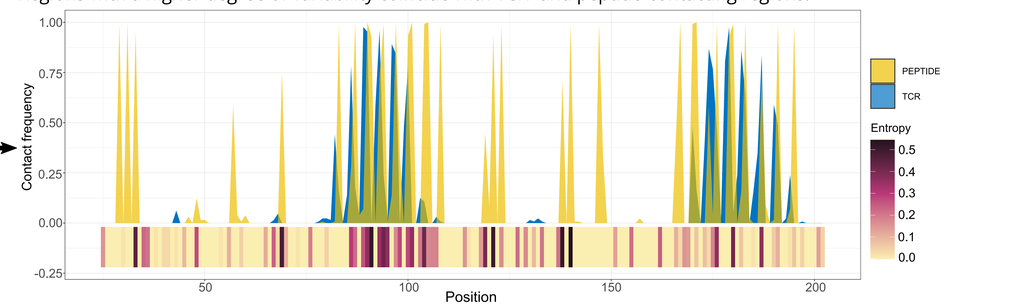


Figure 1. Frequency of contacts with peptide (yellow) and TCR (blue) for different positions of HLA. The populational variability of positions (calculated as Shannon entropy of the corresponding column in MSA) is shown below

2. Variability of TCR- and peptide-contacting positions

Peptide-contacting positions are more variable than TCR-contacting ones.
Most non-contacting positions are highly conservative.

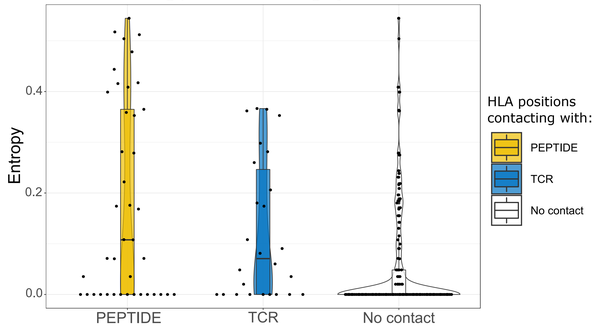


Figure 2a. Variability distribution (Shannon entropy) for peptide-, TCR-contacting and positions which are not in contact with peptide or TCR

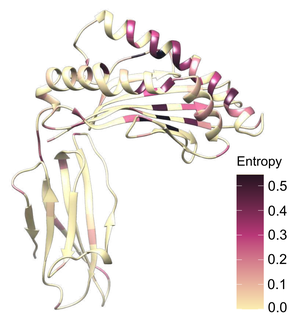


Figure 2b. HLA structure, colored by position variability

3. Sequence variability within HLA-A, HLA-B and HLA-C alleles

HLA-A alleles are more variable in TCR-contacting region compared to HLA-B and -C. While HLA-B are the most variable in peptide-contacting region. HLA-C are less variable compared to HLA-A and HLA-B.

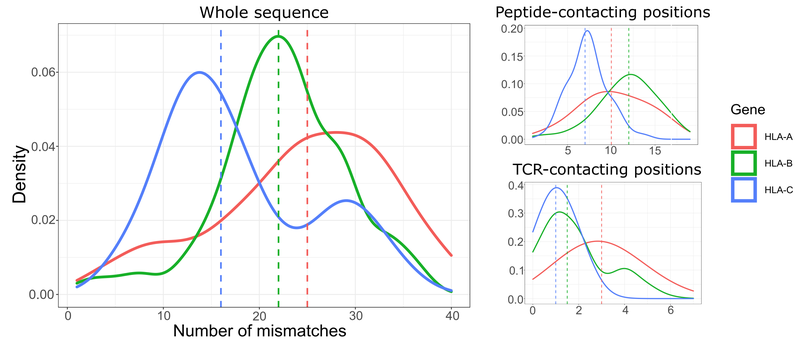


Figure 3. Distribution of the number of mismatches between pairs of HLA alleles of the same gene (HLA-A, HLA-B or HLA-C). Calculated for the whole sequences or only peptide- or TCR-contacting positions

4. Similarity of TCR- and peptide-contacting regions of different alleles

A set of variants with a significant difference in peptide-contacting residues have identical TCR-contacting residues, while the inverse situation is not observed.

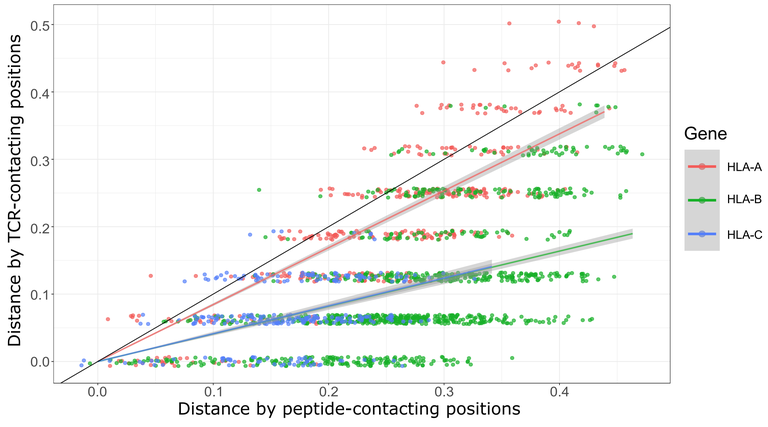
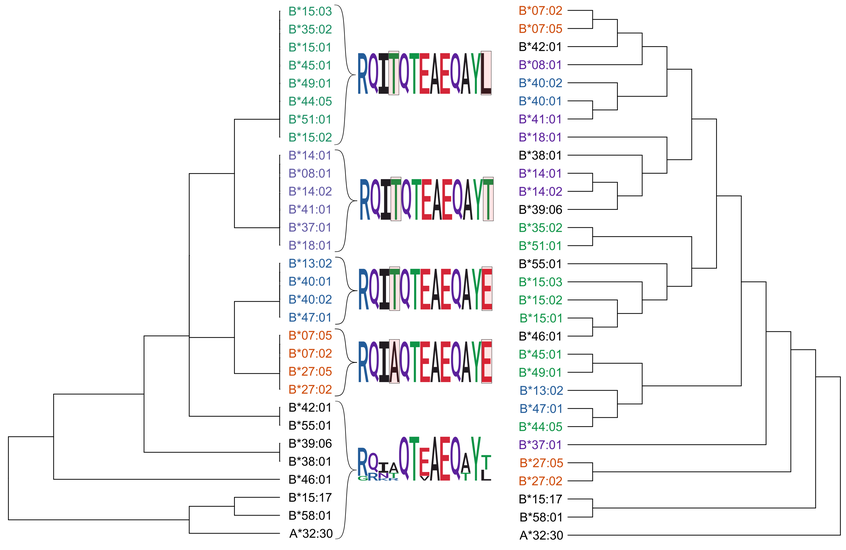


Figure 4. Correlation of pairwise distances between alleles of same HLA gene in TCR- or peptide-contacting regions. Each point represents a pair of alleles, the x coordinate represents difference (number of mismatches) between these alleles in peptide-contacting positions, y coordinate - difference in TCR-contacting positions

5. Evolutionary origin of conservative sequence of TCR-contacting residues



Alleles sharing same TCR-contacting motif are spreaded into different clades on the evolutionary tree, suggesting independent developing of TCR-contacting motif.
The homology of different HLA variants in the TCR contacting region may indicate the optimal structure of this region for interaction with a wide range of TCR variants formed during VDJ recombination.

Figure 5. Dendrograms of HLA-B alleles builded based on the sequence of TCR-contacting residues (left) or the whole sequence (right). Pseudosequences of TCR-contacting region are shown in logo plots