doi 10.18699/bgrs2024-4.1-19

Large-scale analysis of haplotype effects in TCR alpha and beta germlines based on Rep-seq data

Vinogradova S.[1#], Bagrova O.[2#], Vlasova E.[3, 4], Shugay M.[3]*

[1] *Moscow Center for Advanced Studies, Moscow, Russia*

[2] *Lomonosov Moscow State University, Moscow, Russia*

[3] *Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia*

[4] *ITMO University, St. Petersburg, Russia*

[#] *These authors contributed equally to this work*

* *mikhail.shugay@gmail.com*

Key words: T cell repertoires; rare haplotypes; TRA and TRB deletions; V and J gene usage distribution

Motivation and Aim: T cells are an integral part of the human adaptive immune system, with their specificity derived from the unique T cell receptors (TCRs). TCR formation involves a complex process called V(D)J recombination. Through this mechanism, specific V, D, and J genes are chosen from the TRA (TCR alpha) and TRB (TCR beta) loci and are then rearranged to form the TCRα and TCRβ chains of the receptor, respectively [1]. The resulting TCR diversity enables the recognition of a wide array of antigens, facilitating an effective adaptive immune response. To analyse this diversity and track immune responses, a technique called Repertoire sequencing (Rep-seq) is utilised, which involves sequencing of both TCRs and BCRs (B cell receptors) [2]. Recent studies have revealed the presence of both amplifications and deletions within the TRA and TRB genes, as well as in the IGHV loci of BCRs [3, 4]. However, the coherence of these genomic alterations and their biological implications remain unexplored. In this project, we aimed to systematically analyse the TCR repertoires of two large cohorts from Russian [5] and American [6] populations. Our goal was to identify common and rare haplotypes by analysing the frequency of usage of specific V and J genes within individual T cell repertoires. This research not only provides valuable insights into the diversity of the TCR repertoire across these populations but also seeks to explain the observed genomic differences by considering factors such as ethnicity variations across the population, MHC context or thymic selection.

Methods and Algorithms: The study involved two cohorts of donors: Russian donors, consisting of 237 samples subjected to 5’RACE-sequencing, and American donors, with 786 samples obtained from Adaptive Biotechnologies [6]. Analysis was conducted separately within each cohort. Batch effect correction was applied within each group. We normalised gene usage (fraction of reads mapping to a given gene in a given sample) across samples, first standardising to a normal distribution and then scaling back to a [0, 1] interval. Adjusted clonotype frequencies were computed based on these scaled values. We further constructed V/J usage matrices indicating the usage of each segment in samples [5].

Results: We began by analysing the distribution of gene usage across the TCRα and TCRβ chains. While the majority of genes displayed expected unimodal distributions, some revealed bimodal or trimodal patterns. We suggest the bi- or tri- modal distributions are the sign of haplotype effects in populations, appearing due to gene

deletion or homo-/hetero- zygote gene condition. Investigating whether distribution modality could be explained by donor's race, we found statistically significant differences in gene usage between Asian and Caucasian populations for TRBV7-4 gene, and between Latino and Caucasian populations for TRBV4-3 gene. Hierarchical clustering identified rare haplotypes within each population, e.g. TRAJ21, with zero usage in most donors but high usage in a few individuals. Additionally, hierarchical clustering among TRBJ genes revealed two distinct subgroups (J1 and J2) based on usage patterns, possibly influenced by sequence similarity. We further differentiated between functional and non-functional CDR3 sequences, identifying non-functional sequences by stop codons or frame shifts within the CDR3 region. We observed that certain J genes showed a higher usage of non-functional sequences compared to functional ones. While some of these genes are recognised pseudogenes, others without known non-functionality displayed similar patterns, suggesting negative thymic selection against such CDR3 configurations.

Conclusion: This study systematically analysed TCR repertoires in large cohorts from Russian and American populations, revealing significant variations in V and J gene usage across TCRα and TCRβ chains, potentially influenced by ethnicity. Hierarchical clustering uncovered rare haplotypes, indicating unique immune repertoire characteristics within subgroups of donors. Additionally, our investigation into functional versus non-functional CDR3 sequences suggested certain V genes associated with higher frequency non-functional sequences may undergo negative thymic selection, implying a possible biological mechanism where certain gene configurations are selectively disadvantageous. Such T cell repertoire analysis is crucial for practical applications, as understanding the nuances of adaptive immunity can directly influence the selection of targets for receptor-based therapies.

Funding: This work was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation (075-15-2019-1789).

References

1. Murphy K., Weaver C. Janeway's Immunobiology. Garland Science, 2016. doi 10.1201/9781315533247
2. Benichou J., Ben-Hamo R., Louzoun Y., Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology.* 2012;135(3):183-191. doi 10.1111/j.1365-2567.2011.03527.x
3. Omer A., Peres A., Rodriguez O.L. et al. T cell receptor beta germline variability is revealed by inference from repertoire data. *Genome Med.* 2022;14(1):2. doi 10.1186/s13073-021-01008-4
4. Kidd M.J., Chen Z., Wang Y. et al. The Inference of Phased Haplotypes for the Immunoglobulin H Chain V Region Gene Loci by Analysis of VDJ Gene Rearrangements. *J Immunol.* 2012;188(3):1333-1340. doi 10.4049/jimmunol.1102097
5. Vlasova E.K., Nekrasova A.I., Komkov A.Y. et al. Robust detection of SARS-CoV-2 exposure in the population using T-cell repertoire profiling. *BioRxiv.* 2023. doi 10.1101/2023.11.08.566227
6. Emerson R.O., DeWitt W.S., Vignali M. et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet.* 2017;49(5):659-665. doi 10.1038/ng.3822