

# T-cell receptor diversification and specialisation: analysis of $\beta$ -cell repertoire and phenotypic characterisation in response to cytomegalovirus

Popov Ilia

## Abstract

T-lymphocytes differentiated from haematopoietic stem cells are key elements of the adaptive immune response. These cells contribute to immune-mediated cell death through two subclasses: CD4+ and CD8+, which function as "helpers" and "killers" (cytotoxic), respectively<sup>1</sup>. The uniqueness of T-lymphocytes is due to the presence of specific T-cell receptors (TCR) localised on their surface. In the process of T-cell differentiation in the thymus, somatic rearrangement of V, D and J genes belonging to the TR locus genome occurs. The TCR complementarity-determining regions (CDRs) and immunoglobulin (antibody) variable chains are the most variable regions critical for the diversity of antigenic specificities generated by lymphocytes<sup>2</sup>. Divergence of these regions can be assessed by next generation sequencing (NGS) of adaptive immune receptor rearrangements (RepSeq) expressed in the lymphocyte population<sup>3,4</sup>.

**Keywords:** TCR repertoire, T-cell subsets, CMV immunity

## Introduction

Within this study, a total of 16 T-cell receptor  $\beta$ -chain repertoire samples from two individual donors, one of whom was seropositive for cytomegalovirus (CMV+), were analysed. The TCR  $\beta$ -chain repertoire was subjected to sequencing using the 5'-RACE method, revealing a diversity of antigen-specific receptors. T-lymphocytes were classified according to their CD4+ or CD8+ subclasses and phenotypic characteristics, either naive or memory. Two independent blood samples were obtained for each donor/population/phenotype combination, ensuring repeatability and reliability of the data. Clonotypic frequencies were determined by analysing 10,000 sequenced reads for each of the samples, facilitating accurate quantification of the TCR repertoire<sup>5</sup>.

The aim of the present study is to reconstruct sample metadata and identify duplicate combinations. This involves annotating samples (s1 – s16), labelling the corresponding donor, population, and phenotype where applicable.

## Materials and Methods

To analyse the heterogeneity of the T-cell receptor repertoire and annotation results, a complex sample metadata recovery scheme was applied<sup>6</sup>. This scheme involves systematic identification and labelling of samples (s1 – s16), matching them to relevant donors, populations, and phenotypes, and identifying and eliminating potential duplicates in the data.

- To determine the T-cell population, a heat map of CDR3 utilisation was used to categorise the samples into two distinct groups.
- Phenotypic characterisation was performed by analysing correlation patterns between samples.
- Donor samples were identified using a correlation heat map.
- To determine CMV status, an analytical work-up was performed using an antigenic profile histogram.

## Results

### T-cell populations

Groups in Figure 1 show differences in CDR3 utilisation profile, indicating that they belong to different T-cell subclasses. Repertoire annotation performed using the CD8 TRB sequence database provided a higher level of annotation for one of the groups. In addition, pairs of samples were identified and grouped based on similarities in the CDR3 utilisation heatmap, which may facilitate the identification of overlapping replicates in the data.

### Phenotypic characterisation

Figure 2 revealed four pairs of samples with high positive correlation, indicating a specialised response of these cells to a particular antigen. In contrast, the remaining pairs of samples showed significantly lower levels of correlation in the TCR  $\beta$ -chain repertoire, which may indicate their relative naivety or lack of specialisation.

### Donor samples

Analysis of the dendrogram on Figure 2 showed that T cell subtypes form four clusters that divide the axes in half. Given that clustering is not related to population or phenotype, we can assume that samples within each cluster belong to a single donor.

### CMV status

Figure 3 demonstrated that the s3 and s7 CD8 samples belonged to a CMV positive donor. In addition, samples from donors without CMV were found to contain sequences capable of recognising CMV in their CDR3 repertoire. Analysis of the HLA allele histogram on Figure 4 revealed a group of samples with the B\*07 HLA allele, indicating that they belonged to a single donor confirmed as CMV+.

## Discussion

This study successfully reconstructed metadata for 16 T-cell receptor  $\beta$ -chain repertoire samples. The analysis

revealed distinct CDR3 utilization patterns between CD4+ and CD8+ T-cells, enabling population classification. Phenotypic characterization identified potentially antigen-experienced (memory) and naive T-cell subsets. Interestingly, CMV-specific sequences were detected in both CMV+ and CMV- donors, suggesting potential cross-reactivity or exposure history. However, further investigation is needed to definitively assign CMV status to all samples, particularly for CD4+ cells.

Supplementary materials

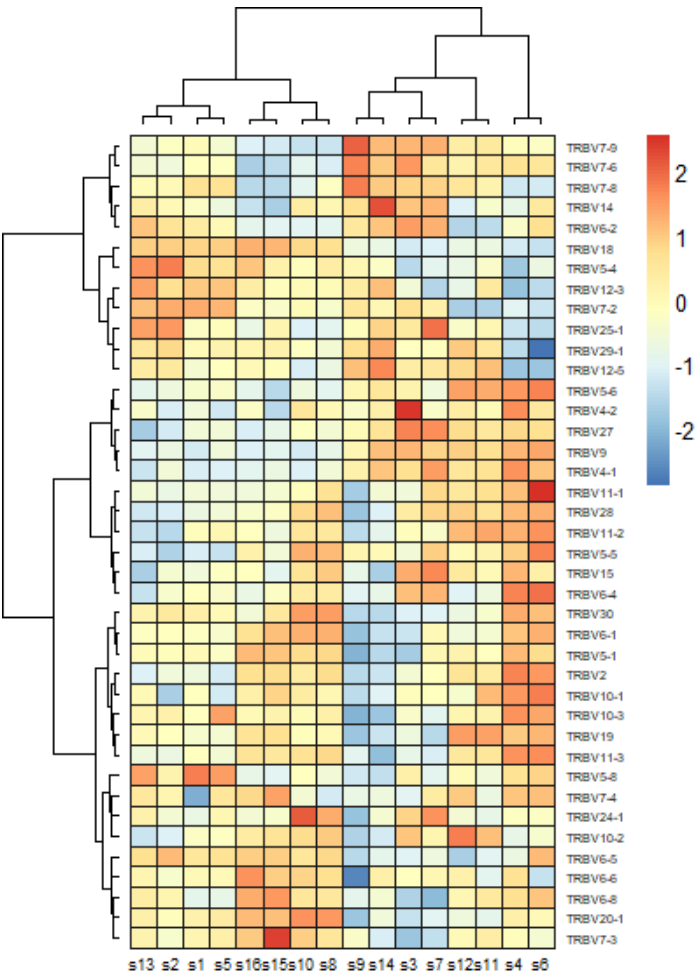
Lab journal can be found in [GitHub repository](#).

Literature cited

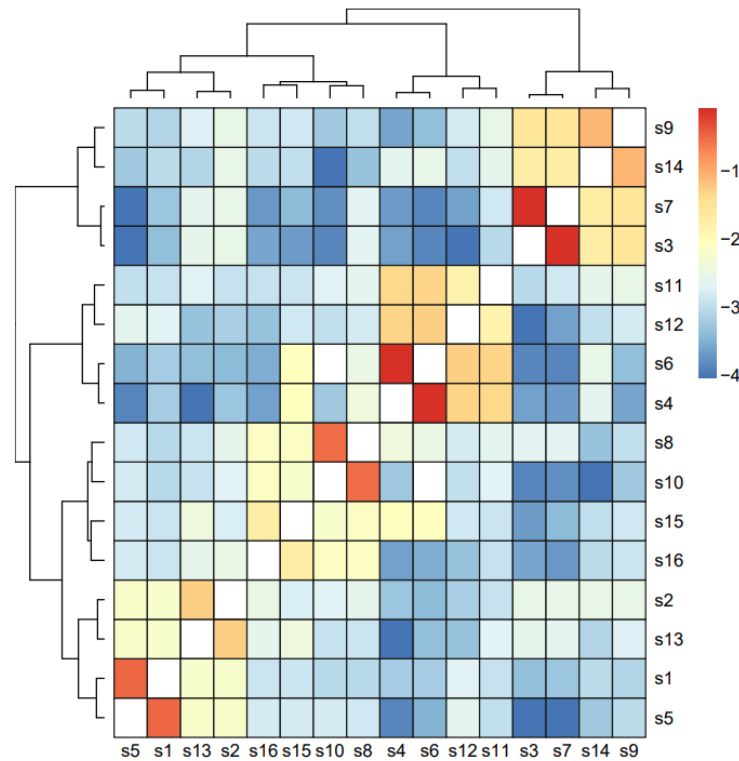
1. Guo, R. *et al.* Generation and clinical potential of functional T lymphocytes from gene-edited pluripotent stem cells. *Exp. Hematol. Oncol.* **11**, 27 (2022).
2. Sun, L., Su, Y., Jiao, A., Wang, X. & Zhang, B. T cells in health and disease. *Signal Transduct. Target. Ther.* **8**, 1–50 (2023).
3. Hou, D., Chen, C., Seely, E. J., Chen, S. & Song, Y. High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease. *Front. Immunol.* **7**, 336 (2016).
4. Chaara, W. *et al.* RepSeq Data Representativeness and Robustness Assessment by Shannon Entropy. *Front. Immunol.* **9**, 1038 (2018).
5. Qi, Q. *et al.* Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci.* **111**, 13139–13144 (2014).
6. antigenomics/repseq-annotation-tutorial: [Tutorial] RepSeq data mining basics in R. <https://github.com/antigenomics/repseq-annotation-tutorial>.

**Table 1** Metadata Summary for TCR  $\beta$ -Chain Repertoire Samples. Asterisks indicate potential donor mismatches for CD4 samples, necessitating a swap of CMV statuses due to the inability to conclusively match CD4 and CD8 cells from the same donor.

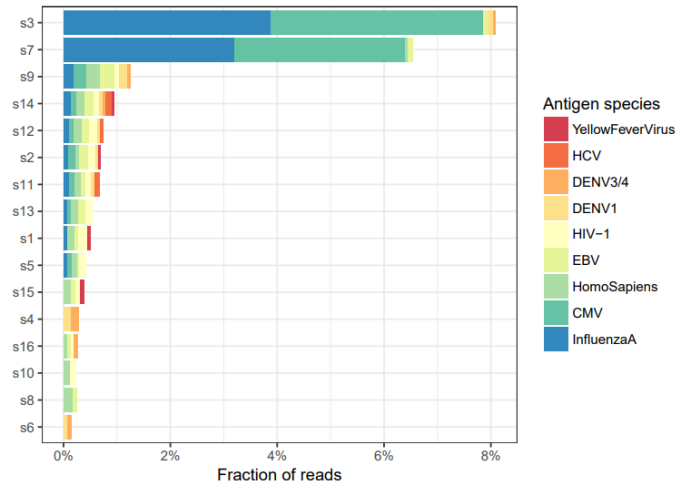
Sample	Donor	Subset	Phenotype	CMV
s3	D1	CD	memory	+
s7	D1	CD	memory	+
s4	D2	CD	memory	-
s6	D2	CD	memory	-
s9	D1	CD	naive	+
s14	D1	CD	naive	+
s11	D2	CD	naive	-
s12	D2	CD	naive	-
s8	D1*	CD	memory	+
s10	D1*	CD	memory	+
s1	D2*	CD	memory	-
s5	D2*	CD	memory	-
s15	D1*	CD	naive	+
s16	D1*	CD	naive	+
s2	D2*	CD	naive	-
s13	D2*	CD	naive	-



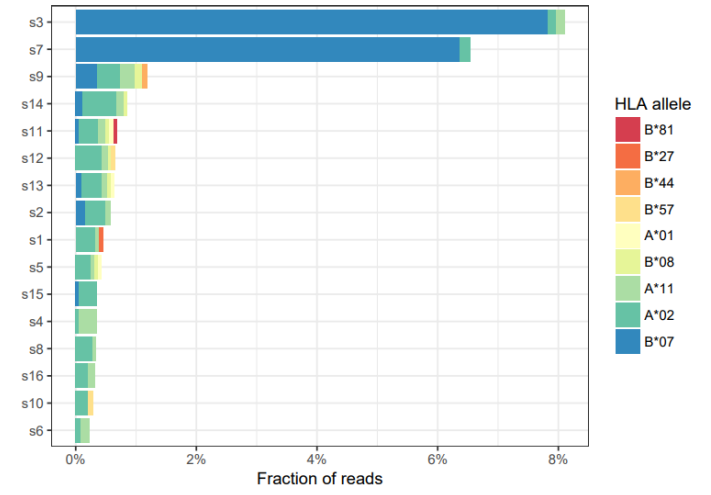
**Figure 1** Heat map of CDR3 utilisation.



**Figure 2** Correlation patterns between samples.



**Figure 3** Antigenic profile histogram.



**Figure 4** HLA allele profile histogram.