Skoltech

Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Artem Mikelov

PhD Program: Life Sciences

Title of Thesis: Dynamics of immunoglobulin repertoires in memory and antibody-secreting B cell subsets in health and disease

Supervisor: Associate Professor Dmitriy Chudakov

Name of the Reviewer:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 05-04-2023**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

The thesis of Artem Mikelov is aimed to study in deep the repertoires of B-cell receptors (BCRs) of different B-cell populations including memory B-cells, plasmablasts, and plasmatic cells, and compare them to naïve B-cell repertoire. Although the plasmablasts and plasmatic cells are extremely rare in the blood and thus the number of the identified clonotypes was limited Artem Mikelov was able to perform a thorough analysis of the data and draw several interesting conclusions. The thesis consists of several sections including Methodology, Results, and a literature review with a list of references. The data is complemented by a sufficient number of figures and tables. The literature review section describes principles of BCR formation, B-cell development, and modern methodological approaches used to study BCR repertoires. The methodological section contains a detailed description of the methods used in the study as well as information about the study cohort and samples. The overall quality of the thesis is high it is well-written and organized in a manner that makes it easy to follow.

The content of the thesis totally corresponds to the topic of the dissertation. The methods used for cell sorting library preparation and data analysis are totally relevant to the aims of the study. The set of surface markers used for the FACS-based isolation of different B-cell subsets is adequate. Next-generation sequencing (NGS) is the most suitable and robust method to characterize adaptive immune repertoires. BCR sequences are very different from other types of genomic or transcriptomic data as the receptors are formed by unique VDJ recombination events in each cell and do not have an exact reference sequence. Thus, there is a number of specific software tools that are used for the analysis of this type of data and Artem successfully applied them to his project. Moreover, he developed his own original approach for the inference of new allelic variants of V-genes that is very important for the correct identification of somatic hypermutations (SHM)

Comparative analysis of the above-mentioned B-cell subtypes repertoires is described in the results section. Additionally, repertoire dynamics is tracked over one year period. Comparative analysis of the cell subsets revealed significant differences in IGH isotype distribution, rate of SHM, and CDR3 length. Correct identification of SHM in BCR sequences is a very challenging task. Germline V-genes have similar sequences and in addition, are quite polymorphic in the human population. One of the main achievements of the current project is the development of MiStrainer - a novel algorithm for V and J gene allelic inference from NGS data. The new algorithm was compared to the existing ones and showed higher sensitivity. Another interesting finding is the increased number of memory B-cell clonotypes shared between individuals indicating convergent selection of BCRs in response to common antigens. This finding is quite surprising taking into account the extremely high potential diversity of immunoglobulins coming from the nature of VDJ recombination. Overall scientific quality of the thesis is high and will have a considerable contribution to the development of the field.

The major results of the current study are published in one of the prestigious international journal (eLife). Artem Mikelov is co-authored two more publications (one of them as a first author) and also reported his results at two conferences. Thus the level of publications related to the results of the current study is satisfactory.

Questions/major comments

1) The threshold for allele inference in MiStrainer algorithm is indicated as 0.35. Each individual genome can normally bear two alleles of the same gene. However, I did not find this restriction in the algorithm. Is this taken into account? What happens when more than two alleles cross the threshold?

2) Why naïve B-cells fractions were not collected from the same donors? Instead, naïve BCR repertoire from another study was taken. Although the repertoire of naïve cells is rather driven by recombination

and should not be very different between individuals of different genetic background selection of V genes could be biased as it was recently shown for T cell receptors (see Corcoran et al., 2023 PMID: 36796364)

3) Why only one replicate was collected for T1? Also, it stated in the methods section that "2 replicates were collected for T2 and T3", however in the table for some PLs only one number is presented. Is it one replicate? This should be clarified in the text

4) Is 1000-2000 cells enough for the reliable estimation of V gene usage? These 1-2K cells (and even less for some samples) will be distributed among 45 V genes meaning only a few clones represent some of them. It could result in unconfident values and huge variability between individuals. The indicated threshold is 2 clones per V gene, but this is very low.

5) "We used repertoires containing more than 5,000 clonotypes and processed them in the same way as our data" – how many repertoires from Gidoni et al were finally taken? I could not find this in the text

6) Number of sequencing reads for UMI – threshold. In the results: "cDNA molecules, each covered by at least three sequencing reads" in the methods "For further analysis, we used sequences covered by at least two sequencing reads". Which one is the correct one?

7) In Figure 6A one of the bars represents bulk PBMCs. However, I did not find bulk PBMC in the data description. What is it?

8) Fig 6B – why different numbers of bars for each individual (from 1 to 3)?

9) SHMs were calculated in general and expressed as an average number of mutations per 100 bp. In my opinion, it can be studied deeper. If it is per bp then we do not differentiate between sense and non-sense mutations that can be significantly biased between B-cell types. The selection should be driven by aminoacid in this sense only the sense mutations matter. And it would be interesting to see their distribution along the IGH sequence to compare different frameworks and CDRs. Although the number of clonotypes is probably not enough to do that.

10) "Of note, the extent of clonal overlap was significantly higher between naive repertoires than for in silico-generated repertoires, indicating functional convergence even in pre-immune repertoires." Not necessarily. The naïve repertoire can be altered by the selection that maybe not be included to the in silico pipeline

11) "PBLs had significantly longer CDR3 regions compared to Bmem cells on average in every isotype except for IgE." – For IgE the difference was not significant most probably due to the very low number of clonotypes. I would stress it in the text

Errors/typos/minor points

Headings of two tables contain "Table 1"

"Before clonal group assignment, we excluded all clonotypes with counts equal to 1". What is the clonotype count? UMI? It is not specified

"800 (623–1,183, n = 9) per 1 x 103 plasma cells" How can it be 1183 clonotypes per 1000 cells?

Fig 6B. Different numbering of Y axis

P57 – a wrong reference to figures: Fig 10 and 11 should be fig 11 and 12

Probably there is an error on Figure 14. The Y-axis is a number of shared clonotypes and each dot should be an integer. Indeed for Bmem random, naïve, and in silico it is the case (0,1,2 or 5), however, for Bmem top it is not

Different numbers formatting through the text (ie. 100 000 and 100000)

Provisional Recommendation

☑ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*