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: Assistant Professor Ekaterina Khrameeva

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 27-03-2023**

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

- **Brief evaluation of the thesis quality and overall structure of the dissertation.**
  The thesis presents an excellent work characterizing BCR repertoires of the three main antigen-experienced B cell subsets — memory B cells, plasmablasts, and plasma cells. The study is longitudal and explores repertoires at three time points over the course of a year. Introduction (Chapter 1), Literature review (Chapter 2), and Methodology (Chapter 3) are written clearly and concisely. The results describing repertoires are presented in Charter 4.1-4.3. In addition, the author developed a new tool for inferring allelic variants of V- and J-genes from such repertoires (Chapter 4.4). This tool is more sensitive than the existing ones, as shown by benchmarking (Chapter 4.5).

- **The relevance of the topic of dissertation work to its actual content.**
  The topic of the thesis matches its contents well.

- **The relevance of the methods used in the dissertation.**

Methods used in the thesis are relevant and applied correctly, except for a few suggestions on possible improvements of statistical analyses, which are detailed below. The used methods are well described and presented with enough details.

- **The scientific significance of the results obtained and their compliance with the international level and current state of the art.**
  The presented research employs advanced experimental and data analysis methods, therefore coping with the international level. Many studies explore B-cell receptor clonal repertoires in disease and control. Yet, the detailed characterization of BCR clonal repertoires in dynamics is still missing, and this thesis fills the gap by describing immunoglobulin heavy chain repertoires in memory B cells, plasmablasts, and plasma cells in dynamics.

- **The relevance of the obtained results to applications (if applicable).**

- **The quality of publications.**
  High enough to pass the PhD program requirements.

The summary of issues to be addressed before/during the thesis defense.

The thesis is clearly written and I have few critical comments regarding its content and presentation of the results. The major concern is the lack of statistical analysis, which could support the conclusions presented in the thesis. For example, there is a statement "The proportion of overall clonal diversity occupied by the five major IGH isotypes was strikingly different between Bmem cells and antibody-secreting cells (ASCs; i.e., PBL and PL)." at page 48. I guess, it related to Fig. 6A. This statement should be supported by formal statistical testing. But I could not find an analysis of the statistical significance of differences presented in Fig. 6A. P-values should be calculated. Fig. 6B is not referenced in the text but the same comment is applicable to this figure. The author should demonstrate the statistical significance of differences between groups. Also, it is a good practice to present the number of observations in groups (or show actual data points instead of bars with whiskers), for better demonstration of initial data supporting the conclusions.

Another potential issue is regarding Figs. 14 and 16. I am not sure it is correct to compare groups of such different sizes (N=15 vs N=4950). Probably, the author should have performed downsampling of larger groups before the statistical testing. Similarly, in Fig. 15, the difference between medians is very minor (~10%) but the groups are very different in size (N>620000 vs N=563). Could the statistical significance be explained by the huge number of observations in the first group?

Regarding the presented novel algorithm, its step 6 (see page 66) is based on the threshold 0.35. It is unclear why this particular value was chosen.

Did the author perform both experimental and data analysis part of the thesis himself? The Results section starts by "We collected…", suggesting that the presented work is a collective effort. So the exact contribution of the author should be clarified.

Minor comments:

- Why is the title of the section 4.5 highlighted?
- Charted 5 is titled "Discussion and recommendations" but it contains only 'Conclusions' paragraph within it. Probably, the chapter title should be reformulated.
- The thesis contains few typos and grammar issues. Still, the remaining typos have to be

corrected.