Prof. dr hab. Anna Żaczek
Laboratory of Translational Oncology
Intercollegiate Faculty of Biotechnology
Gdańsk University and Medical University of Gdańsk
ul. Dębinki 1, 80-211 Gdańsk

Gdańsk 21.02.2023

Review of doctoral dissertation in technical sciences

Discipline: biomedical engineering

Effects of low dose radiation exposure and ageing on T cell receptor beta chain (TCRβ) repertoire in human and mice

Justyna Mika

PhD dissertation of Justyna Mika MSc entitled "Effects of low dose radiation exposure and ageing on T cell receptor beta chain (TCRβ) repertoire in human and mice" performed under supervision of Prof. Joanna Polańska, PhD, DSc and Serge Candeias, PhD, HDR, provides an extensive and unique analysis of T cell receptor repertoire diversity in the context of ageing and ionizing radiation. Powerful analytical pipeline to investigate the TCR diversity based on high-throughput sequencing was developed.

The hardcover dissertation submitted for review is written in English. It contains 163 typed pages and has the layout untypical for doctoral dissertations. It consist of 10 main sections, including typical ones like Introduction, Biological background, Materials and Methods, Summary and conclusions and Supplementary Materials and 5 chapters dedicated to analysis of different datasets, each of them containing separate Introduction, Materials and methods, Results and Conclusions and Discussion sections. Thesis also contains Abstract both in Polish and English, list of abbreviations and a list of almost 200 references.

Untypically, **Introduction section** presents motivation, the aim of the study and the content of dissertation, while state of the art in the field, usually expected in Introduction, is presented in Biological background section.

The work undertakes an important issue of immune system and different factors including sex and ageing affecting its function. Particularly interesting and important from practical point of view is the issue of low-dose radiation impact on the immune system. In the context of presented data on critical role of immune system in health and disease issues raised in the dissertation which have not been described so far, are up-to-date and important to broaden our knowledge in the field. The purpose of the thesis is well defined. It aims to provide an analytical pipeline to investigate the properties of TCR repertoire based on high-throughput sequencing and flow cytometry data. Further the pipeline is to be used to investigate the impact of radiation and aging on various parameters of TCR repertoire, including the diversity of unique rearrangements or the proportion of productive to non-productive TCR sequences.

Biological background section provides biological basis of the project. It presents general information about immune response with a focus on an adaptive subset of immune system. Details about T cells are presented as well as their classification into functional subsets and the process of T cell receptors (TCR) creation by V(D)J recombination. It is compactly written and well-illustrated and still providing very good introduction for further parts of the thesis.

Third chapter entitled **Materials and methods** provides information about databases, datasets and methods used for TCR repertoire examination. Major methods used for TCR repertoire examination are listed, but description of single cell sequencing methods is missing, in my opinion. In this section PhD Candidate describes, for the first time, datasets and methods used for their analysis. Multiple datasets were examined, including data from irradiated mice, healthy donors, and atomic bomb survivors in Hiroshima and Nagasaki. The best approaches were developed, selected and described based on the data preprocessing techniques, statistical and machine learning methods. Also a whole range of methods developed for TCR sequence annotation in short-read sequencing was applied to long-read sequencing data and their performance was investigated. Four hypotheses regarding: i) application of empirical probability models of nucleotide insertions; ii) supporting classical statistical methods by effect size for the big data; iii) application of machine learning to construct different explanatory models; iv) use of non-targeted long read RNA sequencing data to identify V and J gene diversity were clearly formulated.

Next five chapters of PhD dissertation are dedicated to undertaken issues which were as follows:

- Germline DNA retention during the V(D)J recombination process;
- Low dose radiation impact on TCR repertoire diversity;
- Impact of age and sex on TCR repertoire diversity;
- Impact of radiation and age on V gene diversity based on atomic bomb survivors data;
- Analysis of identified TCR sequences from long-reads.

Each of these chapters contains separate introduction, detailed description of datasets and methods, results and discussion. Results are presented very clearly and illustrated with well-prepared figures and tables. Discussion section of each of those 5 chapters relates obtained results to published literature data. In depth discussion of presented issues confirms very good preparation of PhD Candidate and her extensive knowledge in the field.

I find particularly valuable part addressing the question of the impact of radiation on the dynamics of TCR repertoire that has not been investigated before. The radiation effects were estimated using data from atomic bomb survivors in Hiroshima and Nagasaki and data from irradiated mice. Different parameters for TCR diversity were measured: to describe distribution of unique TCR clones (Sequence Diversity) and their number (richness), distribution of unique V genes (V Gene Diversity) and the proportion of productive and non-productive rearrangements (Status diversity) and used to describe the impact of ageing and radiation dose on immune system. Interestingly, it was found out that low-dose radiation accelerates the aging of immune system measured as TCR repertoire in mice. For atomic bomb survivors dataset V gene diversity (13 unique Vβ gene segments: Vβ2, Vβ3, Vβ5·1, Vβ5·2, Vβ5·3, Vβ6·7, Vβ8, Vβ12, Vβ17, Vβ1, Vβ6.1, Vβ13.1, Vβ14), in naive and memory T cells was assessed. Importantly, as far as I understand, those data were generated by flow cytometry so they present protein expression in contrast to the rest of the data in the dissertation generated by sequencing of DNA and indeed presenting gene level information. I am not sure how entitled it is to use term "gene diversity" in this context. May be it should be specified that here proteins not genes are examined? Analysis of multiple exposure parameters interactions revealed that years since bombing had the most significant impact on V gene diversity in naïve cells, while in memory cells additional information about the age of bombing, radiation dose and their interactions was required for better diversity modeling. Intriguingly, when the impact of single V gene on overall V gene diversity was estimated (One-vs-Rest analysis), out of 13 markers analyzed only two (Vβ6.1 Vβ6·7) were not described with unimodal distribution and furthermore GMM was applied for them. Is there any biological rationale for those genes different behavior?

The final part of the PhD dissertation is **Summary and conclusion section** which summarizes obtained results and presents them in the context of hypotheses formulated in the beginning of the dissertation.
This is done in descriptive way. I think it could be advantageous to additionally present the final conclusions in the form of bullet points. With no doubt PhD Candidate achieved set objectives but it could be more underlined in the summary of the dissertation.

Worth highlighting is very good editorial form of the dissertation, which has been written entirely in English, except of Polish version of abstract. The latter is of very poor quality, unfortunately. Also I appreciate showing the sources of funding of the work and I think it is a very good practice to attach scientific achievements of PhD candidate. Scientific achievements of Justyna Mika are significant; she is already an author of 9 publications and she took part in 4 externally funded projects.

While reading, the following questions came to my mind:

- Why only V genes, not whole TCR repertoire, were studied in atomic bomb survivors dataset? While V gene diversity in naïve and memory T cells was examined in atomic bomb

survivors, it would interesting to look at the comparison of this data to general population data. Is it possible?

- Why tools dedicated for single-cell sequencing data (like mentioned in the work SCIGA, scRepertoire, BTCR, SPANTCR) were not included in the analysis of Oxford Naopore Technology dataset? Could the author comment on the significance of results generated based on dataset obtained from 3 donors pooled together (one sample). Is there any difference in information obtained from DNA and RNA sequencing in the case of TCR rearrangements?
- I would agree with the statement presented in the thesis (p.123) that knowledge gained within the project "may be used in many clinical applications and help create new therapies to fight diseases or cancer". As I am particularly interested in cancer, I kindly ask PhD Candidate to discuss this issue in more details, showing some possible scenarios of such clinical application.

Minor comments

- Poor quality of Polish abstract, using terms not used in Polish literature of the field, like "dyskretne segmenty", "genowe dane przekrojowe i podłużne" resulting probably form direct translation of English version to Polish one.
- Some abbreviations are missing in the list, like GMM, FDR.
- According to dedicated guidelines (MIQE) abbreviation RT-PCR in biological sciences means reverse transcription polymerase chain reaction, not real-time PCR. The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines https://doi.org/10.1373/clinchem.2008.112797 propose that the abbreviation qPCR be used for quantitative real-time PCR and that RT-qPCR be used for reverse transcription–qPCR. Applying the abbreviation RT-PCR to qPCR causes confusion and is inconsistent with its use for conventional (legacy) reverse transcription–PCR. But not all authors adhere to this convention.

The above-mentioned inaccuracies or minor errors do not reduce the scientific value of the dissertation. The objectives of the work have been successfully achieved. And the obtained results have significantly expanded the knowledge about TCR repertoire diversity in the context of sex, ageing and radiation. Some of the results have already been published in highly-ranked journals: Frontiers in Immunology and Cellular and Molecular Life Sciences (*Mika J, Kabacik S, Badie C, Polanska J, Candéias SM. Germline DNA Retention in Murine and Human Rearranged T Cell Receptor Gene Coding Joints: Alternative Recombination Signal Sequences and V(D)J Recombinase Errors. Front Immunol. 2019; Candéias SM, Mika J, Finnon P, Verbiest T, Finnon R, Brown N, Bouffler S, Polanska J, Badie C. Low-dose radiation accelerates aging of the T-cell receptor repertoire in CBA/Ca mice. Cell Mol Life Sci. 2017*)

The presented PhD dissertation is of high scientific level, particularly in terms of bioinformatics analysis. PhD Candidate has well managed with the compilation of an extensive and complex research data, careful data curation, in depth analysis of datasets with modern statistical and bioinformatics methods, including machine learning, interpretation of obtained results and their

graphical presentation. Noteworthy, she used a sequence of consecutive computational techniques, which demonstrated her superior skills and practical experience as a researcher. The above mentioned observations deserve high appreciation and a special distinction for PhD Candidate.

Summing up, I am fully convinced that submitted for review doctoral dissertation constitutes an original solution to a scientific problem, presents Candidate's general theoretical knowledge and understanding in the given discipline as well as the ability to conduct independently scientific research, and thus fully complies with the current conditions set forth in the Act on Academic Degrees and Title and Degrees and Title in the Arts (Art 13 of the Act of 14 March 2003 on academic degrees and title and degrees and title in art; Law on Higher Education and Science, Journal of Laws of 2018, item 1669, as amended).

Taking into account the above I have the honor to submit to the Academic Board for the Discipline of Biomedical Engineering at the Silesian University of Technology in Gliwice a request to admit Justyna Mika to the further steps of the proceedings.

Moreover, due to high scientific level and the uniqueness of the bioinformatics analysis carried out as well as already published part of the results in prestigious journals, I kindly recommend this dissertation to be awarded with distinction.