Dr hab. Małgorzata Oczko-Wojciechowska
Department of Clinical and Molecular Genetics
National Research institute of Oncology, Gliwice Branch
Wybrzeze AK 15, 41-102 Gliwice

Gliwice, 07.07.2023

Review of Katarzyna Sieradzka's, M.Sc., Engr. doctoral dissertation entitled: “Classification of white blood cells based on single-cell sequencing data for biodosimetry purposes.”

Supervisor: prof. Joanna Polańska

Advisor: dr Christophe Badie

Single-cell sequencing analysis is used in many of fields, including, cancer, neurology, cardiology and immunology. It is important to emphasize that this type of technology is one of the key applications in cancer. Single-cell analysis is important and powerful tools to get a deeper insight into intratumor heterogeneity. Single cell DNA and RNA sequencing provides gene sequence variation and gene variation and expression data at the level of individual, single cells while bulk DNA/RNA sequencing generate information on the gene variation or transcriptome as the average of all cells: tumor cells and the tumor microenvironment cells.

Although this technology is more and more often used in research, it is not a routine diagnostic activity. For this reason, single cell analysis is still in development and requires standardization of bioinformatics methods. Therefore, the topic presented in the doctoral dissertation is important and necessary.

The dissertation has a characteristic layout for this type of study, it covers 110 pages with references and contains additional materials (pages: 112-149). The author cites 108 publications well matched to the presented topic.

At the beginning of the review, I would like to draw attention to the biotechnological terminology which was used interchangeably, but has a different meaning. This applies to the terms: genetic profile and gene signature. The results presented in the dissertation refer to RNA sequencing and it refers to the genes expression. Genetic profile refers to characteristic change (variants) in DNA sequence of cell as somatic analysis e.g. variants of *BRAF* gene on melanoma cells or germline line

as hereditary predisposition e.g. pathogenic variants of the *BRCA1/2* genes in predisposition to breast or ovary cancer.

In general assessment from a biotechnological point of view, genetic profile it is not the same as gene signature: genomic profile is rather related to analysis of DNA sequence. Genomic profiling look for mutations or other genetic changes in DNA while gene signature is about the activity (expression) of a specific group of genes in a cell or tissue.

Th other important issue is the letter abbreviations (gene symbol) of human genes should be written with a capital letter and italic, without italic, the symbol means human protein.

I. General assessment of the doctoral dissertation composition.

1. The doctoral dissertation motivation is written correctly and justifies the topic.
2. The Introduction and the publicity available tools and development workflows summarizes the available knowledge well and correctly relates to the topic.
   The single cell bioinformatic analysis includes a few the most important steps: alignment, quality control, normalization, feature selections and dimensionality reductions. At present moment there are et least 10 methods available to obtain a low-dimensional representation for scRNA-seq data, non-linear and neuronal network as well. The doctoral student focused mainly on UMAP method, can I ask to justify the choice and what is doctoral student opinion on other methods?
3. In the “Material” section some information is missing: - Where the blood was collected and where it was proceed (Laboratory) as well as where did the blood come from? Volunteers?
   - The consent of the bioethics committee and the biobank where the material is (was stored) was not given
4. There is no clearly defined chapter "results" which would make it easier to navigate through the dissertation and presented results of comparisons of two methods: linear regression and neural networks.
5. Figures presented in “Additional materials” should be in larger size.
6. The “Discussion” section lacked a broader comparison of the prepared algorithm with the available solutions. First of all, highlighting the unique solutions in model based on logistic regression such as parameter panel that can be adjusted according to the needs of the performed analysis. In my opinion this part represent summarization of the results than widely discussion of the developed algorithms.
7. In the presented doctoral dissertation there are some typos and minor language/editorial errors:

- On page 5 is written “behaviors of the genes” what exactly means: genes expression. On page 32 is written that “Table compares the algorithms” instead “tables show the results of comparison”.
- Spaces after brackets or their absence - page 27,
- Citing publication numbers in separate brackets - page 8
- Words in the graphical representation of the algorithm are both, in capital and lower letters pages 28 - 29
- In abstract in polish there were used a sentence” stworzenie odpowiedniego schematu pracy analizy bioinformatycznej”, it would be better to use “opracowanie odpowiedniego schematu analizy bioinformatycznej”.
- few spaces p. 11.

However, I would also like to emphasize that the above minor errors are not significant from the point of view of the correctness of the study and do not significantly affect the reception of the work.

II. General assessment of the results of the doctoral studies.

This doctoral work has two main goals: find the biological information related to genes responded to radiation and development of automated bioinformatic workflow for high-dimentional data originating from single cell sequencing experiment.

The doctoral student present in this dissertation that the prepared algorithm can correctly select genes characteristic of particular fractions of leukocytes subjected to irradiation. This is confirmed by the obtained result (list of the genes), which indicates that the selected genes are associated with the radiation response known from the available literature. The proposed algorithm allows advanced user have more significant impact on the operation of the necessary steps of the analysis and select only relevant steps for performed analysis. It is worth emphasizing that the student attached a great importance to data quality control, which is an important element of this type of analysis. This allowed for reduction of the substantial of heterogeneity od the analyzed data set. Proposed algorithm was used for selected gene signatures characteristic for ionized white blood cells. However, I do not recommend the conclusion, which says that the prepared tool will be used to evaluate the full genetic profile of irradiated cells in the ex vivo environment because the algorithm was prepared to evaluate the signature of genes and for this purpose may be used.

The doctoral dissertation submitted to me for review, prepared by Katarzyna Sieradzka, is a work that I evaluate positively, it is proof of the author's fluent orientation in designing and conducting research and drawing conclusions. I would also like to add that the doctoral student is the first author of two

multi-author monographs and 8 conferences abstracts and the work was supported by European Union grant AIDA.

This work, as well as the monographs achievements of the PhD student, give me the opportunity to conclude that Katarzyna Sieradzka will continue her scientific work and will pose new research challenges.

In conclusion, I state that the dissertation meets the criteria for doctoral dissertations and I am applying to the Biomedical Engineering Discipline Council of the Silesian Technical University in Gliwice for admission to further stages of the doctoral dissertation.

Podsumowując stwierdzam, że rozprawa pani mgr inż. Katarzyny Sieradzkiej pt.: „Classification of white blood cells based on single-cell sequencing data for biodosimetry purposes" spełnia kryteria stawiane rozprawom doktorskim i zwracam się do Rada Dyscypliny Inżynierii Biomedycznej, Politechniki Śląskiej w Gliwicach z wnioskiem o dopuszczenie do dalszych etapów przewodu doktorskiego.