

prof. dr. ir. Dirk Valkenborg  
Interuniversity Institute for Biostatistics and Statistical Bioinformatics  
Hasselt University  
Agoralaan 1, Building D,  
3590 Diepenbeek,

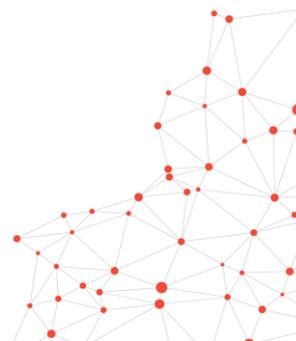
Hasselt, March 9<sup>th</sup>, 2023.

**Review of Bence Galik's Ph.D. thesis** entitled "Development and implementation of original bioinformatics pipelines for NGS data analysis in cancer and assisted reproduction research."

Dear chair of the doctoral evaluation committee,

It was an honor for me to serve as an evaluator of the dissertation of Bence Galik, which contributes to the growing importance of bioinformatics knowledge in biomedical research. The recent scientific approach towards personalized medicine needs a multidisciplinary view of mathematics, statistics, informatics, and molecular biology in a clinical application context. The new generation of biotechnological equipment generates large amounts of raw data requiring high-performance downstream analysis developed by experts with both data science skills and domain backgrounds.

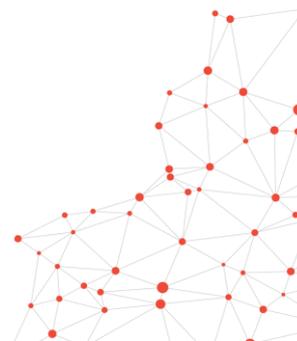
The Ph.D. thesis, "Development and implementation of original bioinformatics pipelines for NGS data analysis in cancer and assisted reproduction research, " demonstrates the crucial role of professional bioinformatics in modern medical science. The thesis covers not only the developed and implemented bioinformatics pipelines applied to various biomedical research



problems, and alone would be more than enough for a thesis, but also the experimental protocols and designs of the wet-lab work. This might also be the ‘weakness’ of the thesis as the material and methods distract from the thorough bioinformatic analysis. The candidate provides five independent experiments as examples of the different areas of bioinformatics and its applications. By their detailed descriptions, Mr. Galik testified his competence as a bioinformatician and biologist, reflected in his co-authorships of peer-review scientific manuscripts.

The first two case studies focus on the mutational landscape of chronic lymphocytic leukemia (CLL) and primary central nervous system lymphoma (PNCLS) in ‘small’ cohorts ( $n = 20$ ,  $n = 77$ ), applying deep targeted sequencing of genes of interest. The downstream analysis of the amplicon data was done using open-source and free well-known bioinformatic tools and data analysis approaches. The candidate pointed out the need for workflow management and reproducibility in the data processing. To exceed this issue, Bence introduced Nextflow as the primary program for this work and the latter case studies. However, as a bioinformatics expert, I would preferably see further detail about the developed pipelines.

In the third study, CpG methylation profiles of sequential glioblastoma samples ( $n = 22$ ) were analyzed by the reduced representation bisulfite sequencing (RRBS) method. The small group size and the difficulty of sample collection and type could create challenges during data analysis. Mr. Galik showed the relevance of finding the best tool for the problem. In the first part of the data analysis, classic methods, like Bismark, were used with some RRBS-specific parameter set (e.g., TrimGalore) to perform methylation calls. In the downstream analysis section, Bence Galik switched platforms and used an R package capable of comprehensive (end-to-end) methylation data analysis obtained with any experimental protocol that provides single-CpG resolution and allows cross-platform analysis. Applying such an approach could maximize the results and gain more information from the data.

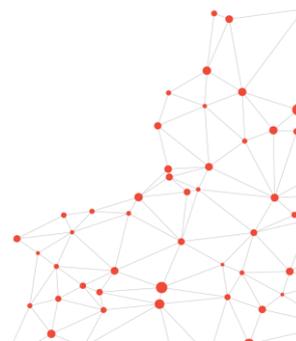


The fourth example is dealing with the miRNAome of non-small cell lung cancer (NSCLC) in a mid-sized cohort ( $n = 177$ ) to understand the disease better and identify potential biomarkers for a more precise diagnosis. The bioinformatics section is well described, especially the quality control steps and their interpretation. This is the only chapter yet to go through peer view processes. Given the promising preliminary results, I strongly support continuing the publishing process.

The last study, which is not cancer-related, focuses on assisted reproduction research. The goal of this study was to demonstrate the use of cell culture media of selected embryos ( $n = 40$ ) to develop a non-invasive copy number variation (CNV) detection method for embryo selection *in vitro* fertilization (IVF) treatments. The low DNA input and low sequencing coverage could make data processing difficult. The embryo selection, wet-lab, and dry-lab methods are described step by step. The data analysis steps depended on widely used bioinformatic tools.

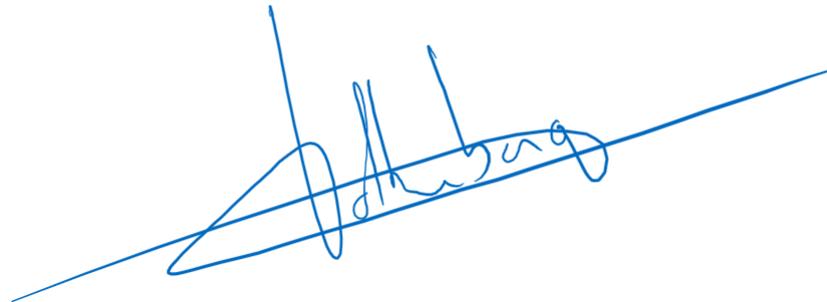
The language of the monograph is of good quality. Still, there are typos (e.g., misspellings, missing letters) and incoherent use of technical terms (p.e.: Kruskal Wallis vs. Kruskal-Wallis). The Latin expressions must be in italics, such as *in silico*, *postmortem*, *in vivo*... Some of the material and methods parts are unclear and need further polishing; a few of the catalog numbers of the reagents are included, while most of them are excluded. For reproducible science, mentioning the provider and catalog numbers of the equipment and consumables used in this dissertation is essential. The reference number of the ethical permissions is also missing in some cases. My last observation about the editing, there are three missing references on pages 49, 98, and 99.

In the last part of the discussion, I recommend adding some extra sentences to underlie the relevance of the bioinformatics work and highlighting that the thesis is mainly about downstream bioinformatics analysis; the experiments conducted during the thesis are tools that provide data for it. Still, the exciting biological findings take away the emphasis from the



original aim of the thesis, which is to represent the value of bioinformatics in modern medical science. The wet-lab and dry-lab sections are not in harmony, and the balance should be tuned to bioinformatics.

The author demonstrated his excellent knowledge of designing, developing, and implementing various data analysis workflows. The submitted thesis meets the requirements of the doctoral regulations of the Medical University of Bialystok to obtain the academic doctoral degree. Thus, I ask the Senate of the Medical University of Bialystok to admit Bence Galik to the further stages of the doctorate process, which may be subjected to public defense.



Yours sincerely,

prof. dr. ir. Dirk Valkenburg

