Review of the PhD thesis of Mr. Geroges Bedran.

The submitted PhD thesis entitle **Exploring alternative sources of tumor antigens using large-scale immunopeptidomics** by Georges Bedran focuses on computational analyses of large-scale publicly available mass spectrometry immunopeptidomics datasets, with the aim of identifying novel sources of peptides that could be potentially important for immune recognition of cancer cells. His main objective was to generate new computational pipelines suitable for the analysis of such MS data.

The thesis starts with a short introduction covering very briefly various topics about HLA-I and HLA-II complexes, the presentation machinery, general concepts about cancer and the immune system, and brief description of methods for identifying mutated tumor antigens by genomics and in silico prediction, mass spectrometry based immunopeptidomics, and an overview of some sources of antigens, particularly, genomic variants, alternative splicing, non-canonical regions, and post translational modifications.

Next, the aims of the thesis are outlined. They are categorized into technical aims and biological aims that address the gaps in our understanding of the non-canonical and post-translational MHC-I immunopeptidomes, and their association with various cancer types for assessing their tumor selectivity. Two manuscripts are then appended. The first describes a new computational pipeline, developed by Georges, for identification of peptides derived from novel ORFs in coding genes (3-frame translation of pre-mRNAs of protein coding genes), resulting in the identification of thousands of peptides from large-scale publicly available datasets, using available de novo sequencing and open search tools that were assembled into a dedicated pipeline. This work was done under the supervision of Prof. Javier Alfaro. The second manuscript describes the implementation of a group specific FDR in elements within the MSFragger environment (developed by the group of Prof. A. Nesvizhskii), allowing for identification of glycosylated peptides with improved accuracy. Investigation of publicly available immunopeptidomics datasets revealed a prevalent presentation of gylco-peptides on HLA-II complexes, with almost not presentation on HLA-I complexes, and a different composition of the glycosylations compared with proteomics data. The association of the gylco-HLA-II peptide with the HLA-II binding pockets was assessed with available motif deconvolution and binding affinity prediction tools. This work was done under the supervision of Prof. Alexey Nesvizhskii. Last, the thesis is ended with a detailed literature review on computational approaches for antigen recognition by prediction of immunogenicity of peptides, structural information, and through TCR sequencing and machine learning approaches.

The presented research in the thesis, included in the two manuscripts is original, of high quality and is very well written. The first manuscript was already peer-reviewed and the second manuscript, which I found to be more innovative, is very timely and will likely be evaluated very positively by the scientific community. The introduction, aims, and the summary, milestones and future directions sections were somehow less coherent. The introduction covers indeed the relevant topics, though it is very short and not detailed enough. For example, interesting classes of very important known tumor antigens were not mentioned at all in the thesis introduction, such as shared tumor antigens, that have been extensively studied by immunopeptidomics and were tested in the clinical as immunotherapy targets. In addition, there is room for providing more elaborate explanations regarding the hurdles associated with identifying HLA peptides in MS data through proteogenomics, as well as the diverse MS acquisition methods and their respective benefits. Additionally, a more comprehensive overview on the significance of controlling the false discovery rate in extensive database searches would have been valuable, to better appreciate the new and elegant computational solutions developed in this thesis. Some key papers are cited but are not discussed. The introduction sections in the two manuscripts are better written and they provide a good overview of the literature.

The summary, milestones and future directions section contained a short discussion on the current work and on the future directions, and it contains a very elaborated literature review on current approaches for prediction of MHC-peptide-TCR recognition. This section is very well written but it is not clear how this content is directly linked with the two main narratives of this thesis. This section could have included an explanation on what are the biological significances of the results and discoveries presented in the thesis, for example, why glyco-peptides were not observed in the class I peptidome. Additionally, it could have explored potential avenues for further development of these approaches and addressed the persisting challenges associated with identifying actionable cancer-specific antigens through MS-based immunopeptidomics. The current content in this section might be more suitable to be published as an independent review manuscript.

The doctoral dissertation meets the requirements set for doctoral dissertations by The Higher Education and Science Act dated 20 July 2018 (Polish Journal od Laws od 2018 item 1668, as amended). The assessment of the work is positive.

Sincerely,

Michal Bassani-Sternberg