Thesis Evaluation

Ph.D. candidate: **Georges Bedran**

Thesis title: **Exploring alternative sources of tumor antigens using large-scale immunopeptidomics**

Understanding of the Subject Matter:
The thesis comprises four chapters. The first chapter is an introduction that provides a presentation of the field of immunopeptidomics, its applications and limitations. The presentation is well articulated and emphasizes the needs for comprehensive bioinformatics tools to identify antigens from all genomic regions. Chapter two presents the analyses of 26 MHC class I immunopeptidomic published studies across 11 different cancer types where identified more than 8000 non canonical MHC I-associated peptides (MAPs) together with modified residues were corroborated using spectral matching and false discovery rate (FDR) ≤ 1%. This study was recently published in the journal Cancer Immunol. Research. Chapter 3 presents a novel computational workflow (HLA-Glyco) to identify glycosylated MAPs from 8 publicly available immunopeptidomic studies where these modifications were mostly represented in MHC class II datasets. This resource contains over 3,400 human leukocyte antigen (HLA) class II N-linked glycopeptides from 1,049 distinct protein glycosylation sites, and revealed a high levels of truncated glycans, and differences in the specificity glycans between classical HLA class II allele groups. This work is posted on Biorxiv and awaits comments from reviewers. The final chapter summarizes key findings observed and provides thoughts on future directions. This thesis presents general theoretical knowledge and original methodology for a Ph.D. student in bioinformatics. Overall, the thesis is of high quality and provides a body of new knowledge on immunopeptidomics.

Contribution of the Thesis to Advancing Knowledge:

The originality and significance of the study reported in this thesis is reflected by the publication of a first author papers in Cancer Immunol. Research (IF: 12) and submission of a separate study in another high impact journal. These key papers provide valuable biological insights into immunopeptidomics and the occurrence of non-canonical MAPs. Bioinformatic tools developed revealed a subset of 8,601 ncMAPs specific and common to different types of cancer. The use of immunopeptidomic data from normal tissues facilitated the development of filters to set threshold for the selection of ncMAPs in healthy tissues.

While tools enabling the identification of glycopeptides from proteomic datasets are currently available, none of these are designed to search efficiently large-scale immunopeptidomic data. The use of MSFragger-Glyco along with the comparison of detected glycosylation sites with the proteome, and the correlation with HLA motifs between glycosylated and non-glycosylated provides a novel strategy to study glycosylated MAPs without any enrichment. This study provides a comprehensive repertoire of glycosylated MAPs and revealed truncated glycoforms and distinct specificity across allele groups.

For both studies described in this thesis, the candidate demonstrated critical thinking and the capability to conduct scientific research independently. This thesis provides an original solution to the identification of glycosylated MAPs and antigens arising from non-canonical genomic regions, which represent challenging problems in the fields of bioinformatics and immunopeptidomics. Overall, this thesis presents remarkable work with tangible outcome for the selection of ncMAPs and putative tumor-specific antigens.

Organization, Writing Style and Presentation of Material:
The thesis is extremely well written and easy to read. The data are convincing and figures for the most part clearly laid out. I wrote minor comments on the margin of the thesis, and will make my copy available during the thesis defense.

Summary:
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).

Revisions Necessary for Thesis to be accepted:

- p.1: 'Under healthy and diseased conditions, these proteins are degraded by the proteasome', please mention UPS degradation pathway. The current thinking is that defective ribosomal products or protein retirees are modified by ubiquitin chains prior to their degradation by the proteasome.

- P. 7, 'Although fast and inexpensive, indirect methods lead to suboptimal neoantigen prioritization owing to the discrepancy between the theoretically possible MHC associated peptides and the experimentally presented ones'. Should indicate that predictive approaches that use RNASeq and HLA binding tools fail to recognized antigen processing steps by amino peptidases, which degrade many of the putative HLA binders.

- P.7, section on direct identification of MAPs by MS, it would be pertinent to reference several reviews on immunopeptidomics (e.g. PMID: 37003057, 36841147, 35716458, 34635837, 34310018, 33424836).

- P. 9, the section on promising sources of antigens would benefit from an introduction on how previous studies integrated proteogenomic approaches (ie Ribosome profiling: PMID: 34663921, cancer-specific proteomes PMID: 35367648, 30518613) for the generation of Databases that are subsequently searched by different engines

- P. 24, 'The non-canonical presentation appeared to be 5 times enriched for the A03 HLA supertype, with a projected population coverage of 54.85%'. Is this statistically significant and why allele A03 would present a higher distribution of ncMAPs?

- P. 36, in the close open de novo section, De novo sequencing software (e.g. DeepNovoV2) can produce a high FDR, and can't necessarily account for peptide isomers that can match the same MS/MS spectrum. Was there any attempt to

determine FDR with a learning dataset or to compare results with another de novo software?

- P. 36, ‘a 3FT database to reveal the de novo-based ncMAPs. Does this step enables genome alignment of putative sequences and how candidates are selected when multiple hits are found?

- P. 37, ptmMAPs, The rationale for the selection of PTMs is not clear. For example, why phosphorylation was not selected?

- P. 37, Figure S2a also disclose that 23.5% of PTMs identified are consistent with original studies. Does this reflect the lack of consistent search parameters or else?

- P. 38, Numbers are higher in Fig 3e compare to 3c. Is there an explanation?

- P. 39, why a different FDR was selected (e.g. 10%)?

- P. 42, Moreover, to ensure low toxicity levels… (Fig. 6c). GTEX filtering of putative TSAs without TPM in normal account for significant depletion of candidates. Wouldn't it be easier to consider a database constructed from cancer-specific specimens?

- P.75, we focused exclusively on N-glycosylated MAPs. This could be an unnecessary limitation. O-GlcNAc peptides also produce oxonium ions that would be identified in MS and could be assigned to Ser or Thr residues present.

- P.78, Overall, the data showed… Was there any evidence of O-GlcNAc with m/z 163 oxonium fragment ions in either HLA I or II?

- P. 101, The perspectives section could describe how deep learning can be used to provide better de novo sequencing and correlation to transcriptomic data for genome annotation. The modeling of TCR-MHC is great but falls outside of the scope of the present thesis.

Recommendation:

The assessment of the thesis is positive, proceed to Defence.

Pierre Thibault

Date: May 30, 2023

Pierre Thibault, Ph.D.

Professor and CAPCA Director
Institute for Research in Immunology and Cancer
Department of Chemistry
Université de Montréal