L-Università ta' Malta

**Faculty of Health Sciences**
**Dept. of Applied Biomedical Science**
Rm 417, Biomedical Sciences Building
University of Malta,
Msida MDS 2080 Malta
T +356 99945592 david.saliba@um.edu.mt

Report for Doctoral Thesis entitled: "Development of computational methods for the high-throughput analysis of immunopeptidomic data and their application to artificial intelligence (AI)-enabled precision medicine". Ashwin Adrian Kallor, M.A.

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)

In this dissertation the candidate presents a novel framework for analysing MHC-peptide promiscuity to improve cancer vaccine development and offers several unique insights into neoantigen sharing across diverse HLA backgrounds. The impressive success of immune checkpoint blockade treatments has inspired scientists to identify immunogenic epitopes and use them in the development of personalized vaccines. Peptides that bind to human leukocyte antigens (HLAs) and originate from the processing and presentation of mutated proteins are among the primary targets recognized by T-cells in cancer cells. HLA-I molecules present neoantigens to CD8+ T cells, enabling immune recognition of tumours. However, not all mutations generate peptides that bind MHC-I effectively are immunogenic. For example, immunogenic neoantigens exhibit stronger HLA-I binding affinity and longer complex stability when compared to non immunogenic neoantigens. Moreover, even highly expressed neoantigens may fail to trigger responses if not efficiently presented by HLA-I. Mass spectrometry-based immunopeptidomics (e.g., detecting MHC-I-presented peptides) is critical for validating candidates. Through the development of the CARMEN framework, the candidate advances these concepts by identifying promiscuous neoantigens that bind multiple HLA-1 alleles thereby increasing the likelihood of presentations across diverse patient populations with a few albeit important exceptions.

In Chapter 1 of the results the candidate presents the methodology for data collection and harmonisation of pan-cancer data from 72 publications (2323 samples) used to engineer the CARMEN database. Most earlier analyses, such as those identifying shared neoantigens for off-the-shelf vaccines, have been limited to a handful of common HLA haplotypes. However, CARMEN provides a robust resource for identifying shared neoantigens across HLA backgrounds. Here, the student could have commented on the strength or limitation of this approach considering the heterogeneity in source studies which might have introduced biases based on varying sequencing depths or HLA typing methods. Comparing the CARMEN approach to the recently published ONMI-MHC methodology (published in 2025 after thesis submission; doi: 10.3389/fimmu.2025.1550252), which trained on standardized mass spectrometry and binding data, would help consolidate the strength of the vast dataset used in this thesis. Nonetheless, Figure 18 and table 5 in this chapter is a testament to the rigour of the depth of information of the CARMEN database. Cross referencing the most frequently recurring non-canonical peptides (Table 5) with ribosome profiling

and ORF features (doi.org/10.1038/s41467-024-46240-9) could provide important insights on the genomic locations of these. The use of Gibbs clustering and UMAP dimensionality reduction, defined by the candidate, enabled the systematic mapping of the motif landscape. This study moves beyond individual-level predictions and provides population-scale insights by analyzing population frequencies for each of the peptide motifs. In the final part of the chapter the candidate describes the use of PrimeCUTR to investigate the relative contributions to neoantigen landscape from different mutation classes with an emphasis on 5'UTR sites. Whilst this remains *in silico* work, it would be interesting to see whether these novel neoantigens are indeed immunogenic. Furthermore, it would be valuable to determine if the function of anchor residues is preserved in these non-canonical peptides. Of interest, is the suppression of start-gain formation following UV exposure in Melanoma samples (Figure 27). The findings presented in this chapter are examined comprehensively in the discussion section, with thorough engagement with seminal literature in the discipline. Furthermore, the chapter elucidates the notion of peptide "promiscuity" in relation to their capacity to bind multiple HLA alleles.

In Chapter 2, the candidate presents a novel framework for analyzing MHC-peptide promiscuity to improve cancer vaccine development. Here the candidate maps MHC-I peptide promiscuity from physical evidence detected by mass spectrometry (MS) through a multi-step integrative approach combining immunopeptidomics data with genomic information. Briefly, detected peptides are systematically aligned back to the human genome to identify their precise genomic origin. This mapping allows the identification of *contiguous genomic regions*—termed **epitope contigs and scaffolds**—which represent clusters or "hotspots" of antigen presentation rather than isolated peptides. This aggregation captures *hotspots* of antigen presentation rather than isolated peptides, reflecting biologically relevant regions that generate multiple presented epitopes. For each peptide and genomic region, the candidate defines promiscuity scores (defined by the Gibb's cluster motif space), unique peptides, expression levels and HLA-I allele association. The results defined in Figure 47 of this thesis clearly depicts the gene hotspots with classical gene targets such as TP53 and KRAS. This methodology moves beyond prediction to provide *empirical, genome-resolved maps* of antigen presentation, enabling more accurate identification of neoantigens shared across diverse patient populations. Finally, an SVM (Support Vector Machine) model was trained using features derived from contig-overlapping mutations, including promiscuity metrics and mutation burden in these regions. The model identified contigs whose mutation status correlated with better immunotherapy response, suggesting these regions are critical for effective anti-tumor immunity. By focusing on promiscuous peptides and contigs presented across many HLA-I alleles, the approach identifies neoantigens likely to be shared among diverse patients, increasing the feasibility of "off-the-shelf" vaccines rather than fully personalized ones.

In conclusion this thesis moves beyond prediction to provide *empirical, genome-resolved maps* of antigen presentation, enabling more accurate identification of neoantigens shared across diverse patient populations. Additionally, this thesis underscores that HLA-I biology is central to both neoantigen immunogenicity and patient-specific immunotherapy outcomes. By systematically analyzing promiscuity, CARMEN provides a roadmap for designing vaccines with broader applicability and refining predictive biomarkers—advancements critical for overcoming resistance in heterogeneous patient populations.

The thesis is characterised by a notably intricate writing style; nevertheless, the principal arguments are articulated through a coherent and methodical progression that facilitates seamless transitions between chapters, thereby augmenting both readability and intellectual engagement. The candidate's ability to interpret research findings within the current field of knowledge not only demonstrates their capability but also propels the boundaries for future investigations. Collectively, this dissertation highlights the candidate's capacity to autonomously conduct scientific research, showcasing a remarkable skill set and expertise in the field.

Therefore, I am applying to the Council of the Biotechnology Discipline for admission of *Mr Ashwin Adrian Kallor* to further stages of the doctoral procedure.

Yours Sincerely,

Prof. David G. Saliba (PhD Edin.)