Georges Bedran

Exploring alternative sources of tumor antigens using large-scale immunopeptidomics

Abstract

The identification of cancer neoantigens is propelling a new era of vaccines and antigen-specific T cell therapies. Mass spectrometry has been the sole high-throughput approach for characterizing the physical presence of neoantigens in cancer. Early efforts to investigate antigen presentation focused on combining publicly available studies to query canonical MHC-associated peptides (MAPs). However, the profiling of non-conventional antigens, such as non-canonical (i.e., translation of non-coding regions) and post-translationally modified MHC-associated peptides, remains limited and is rarely clearly understood.

In Chapter Two, I developed a proteogenomic pipeline based on deep learning *de novo* mass spectrometry to enable the discovery of non-canonical MHC-associated peptides (ncMAPs) from non-coding regions. Considering that the emergence of tumor antigens can also involve post-translational modifications, an open search component was included in the pipeline. Leveraging the wealth of mass spectrometry-based immunopeptidomics, I analyzed 26 MHC class I immunopeptidomic studies of eleven different cancer types. I validated the *de novo* identified ncMAPs, along with the most abundant post-translational modifications, using spectral matching and controlled their false discovery rate (FDR) to 1%. Interestingly, the non-canonical presentation appeared to be 5 times enriched for the A03 HLA supertype, with a projected population coverage of 54.85%. I revealed an atlas of 8,601 ncMAPs with varying levels of cancer selectivity and suggested 17 cancer-selective ncMAPs as attractive targets according to a stringent cutoff.

In Chapter Three, I developed a glyco-immunopeptidomics method using the ultrafast glycopeptide search of MSFragger and several layers of stringent control of false discovery

rates. I performed a harmonized large-scale analysis of eight publicly available studies to produce a resource containing over 3,400 HLA class II glycopeptides from 1,049 distinct protein-glycosylation sites. I revealed characteristics in HLA glycopeptides, including high levels of truncated glycans, conserved HLA-binding cores across the 72 studied HLA class II alleles, and a different glycosylation positional specificity between the classical allele groups. With the goal of supporting further development in the nascent field of glyco-immunopeptidomics, I provided a reproducible glyco-immunopeptidomics pipeline within the fragpipe suite along with a web resource for ease of access.

In Chapter Four, I conclude this thesis with a summary of my findings, a discussion of the unmet needs in the field, and my vision of the research to come.

The establishment of both the non-canonical and glycosylated landscapes of MHC-associated peptides within the framework of my PhD represents a milestone towards understanding the complexity of the immunopeptidome and paves the way for broader therapeutic research against cancer.