Project Proposal QCB

**Integrative Machine Learning Framework for Accurate Prediction of Peptide Immunogenicity**

The accurate prediction of peptide immunogenicity is a critical challenge in immuno-oncology, vaccine design, and infectious disease research. Despite the increasing availability of sequencing data and computational tools for neoantigen discovery, the majority of peptides predicted based solely on MHC binding fail to elicit functional T cell responses. This proposal aims to develop a robust machine learning (ML) framework to predict peptide immunogenicity by integrating multi-dimensional features that go beyond MHC binding affinity, including antigen processing, TCR recognition potential, peptide-MHC stability, sequence motifs, and immune context. Our approach will leverage large-scale datasets of experimentally validated immunogenic and non-immunogenic peptides, encompassing cancer neoepitopes, pathogen-derived antigens, and synthetic peptide libraries. We will implement both deep learning and ensemble ML methods to identify patterns in immunogenic peptides that are not captured by current models, using sequence embeddings, physicochemical descriptors, and structural features derived from AlphaFold2 models.

The model will be trained and evaluated using rigorous cross-validation and tested on independent datasets from public repositories (e.g., CEDAR, TESLA) and in house in vitro T cell activation assays. To enhance interpretability and biological insight, we will incorporate explainable AI methods to identify the most predictive features and assess their contribution to immunogenicity across peptide classes and HLA alleles. By combining cutting-edge machine learning (ML) techniques with immunological expertise and high-throughput data, this project aims to develop a next-generation immunogenicity predictor that surpasses current benchmarks and accelerates translational efforts in cancer and infectious disease immunotherapy.

Further Readings:

* Albert et al. Deep neural networks predict class I major histocompatibility complex epitope presentation and transfer learn neoepitope immunogenicity. *Nature Machine Intelligence* 2023
* Wohlwend et al. Deep learning enhances the prediction of HLA class I-presented CD8+ T cell epitopes in foreign pathogens. *Nature Machine Intelligence* 2025.

Contact: michele.ceccarelli@unina.it