XNeoCrypt: An Interpretable Ranking Framework for Prioritizing Noncanonical Tumor Antigen Candidates

Yu, Aaron (School: Millburn High School)

With over 1.8 million deaths from lung cancer annually, it is imperative to develop immunotherapies that precisely target these tumors while minimizing off-target effects on healthy cells. A major bottleneck is identifying the tumor-specific antigens to target, which many current pipelines fail to solve at scale because they focus solely on DNA mutations, missing a large class of peptides caused by abnormal splicing. Recent studies have shown that these splice junctions between exons and transposable elements (JETs) can generate antigens that are highly recognizable by T cells and are specific to tumors. However, there remains no clear and interpretable way to decide which JETs are worthy of experimental validation. I engineered a novel computational framework to systematically detect and rank these JET-derived antigens in the LLC1 mouse model. The ranking framework uses interpretable, evidence-driven features such as MHC-I binding and chemical properties from each peptide to prioritize high-confidence targets for experimental validation and avoid targeting healthy cells. From over 123,487 peptides screened from a public RNA-seq dataset, I reduced the search space by over 4,900-fold to a top-25 candidate shortlist that successfully prioritized known JET candidates above random baseline across controlled evaluation trials with high statistical significance ($p<0.0001$). Top-ranked peptide-MHC complexes were further analyzed through structural modeling with AlphaFold to support interpretability. This turns a colossal search problem into an actionable set of targets that can move into wet-lab immunopeptidomics or T-cell assays, substantially reducing experimental burden and enabling practical noncanonical antigen discovery.

Awards Won: