AACR abstract

A novel algorithm to identify TCR-binding somatic mutations from human cancers

Priyanka, Malini, Kiran, Ravi, Rohit Amit

A large number of pre-clinical and clinical studies have strongly implicated that combining cancer-derived neoantigen vaccines with checkpoint control inhibitors will enhance priming and expansion of tumor-specific naïve and memory T-cells resulting in superior efficacy and durability of response. The neoantigens, derived from somatic mutations are prime candidates for cancer vaccines, not subjected to host's central and peripheral tolerance. Identifying potential T-cell engaging neoepitopes among a large number of somatic mutations is like finding a specific needle in a stack of needles. Currently, the available T-cell neoepitope prioritization pipelines rely primarily on two attributes - the class-I HLA-binding affinity of the mutant peptide compared to the wild-type counterpart, and the level of expression of the mutated gene in tumor cells. The higher the binding differential and higher the expression level of the mutant allele, the greater is the likelihood for the peptide to be presented on antigen-presenting cells. These approaches, however, fall short of predicting whether the HLA-bound peptide will engage T-cells by binding to T-cell receptors (TCRs). We have developed a novel algorithm to predict the binding of HLA-peptide complexes to TCRs by analyzing the physicochemical composition of the amino acids and their positional biases in the 9-mers from crystal structures of HLA-peptide-TCR complex. We applied machine learning approaches to build a classification model that can predict whether a given 9-mer peptide is a TCR-binder or not, by identifying whether an amino acid at a given position carries key features that will facilitate interaction with the TCR. We applied this approach to positive and negative TCR interactions selected from Immune Epitope Database (IEDB). We tested multiple classification approaches and found that Random Forest and ClassificationviaRegression methods provided the best performance. We achieved more than 99% accuracy at 10-fold cross validation on both training and unseen test datasets. We further validated our model using positive and negative peptides curated from published papers reporting clinical trial results of checkpoint control inhibitors. The performance of the two classification models was evaluated by two different TCR-binding assays - dextramer binding and IFN-γ release. The ClassificationviaRegression method showed a higher positive predictive value for the dextramer-binding assay, whereas the Random Forest method showed a higher positive predictive value for the IFN-γ ELISPOT assay, suggesting subtle differences between the two classification methods. The inclusion of the TCR binding step to our T-cell neoepitope prioritization pipeline increased the accuracy of prediction, reduced false positives and selected potential neoepitopes to a manageable number for testing in cell-based assays.