nature biotechnology

The problem with neoantigen prediction

Personalized immunotherapy is all the rage, but neoantigen discovery and validation remains a daunting problem.

ast December, the newly minted Parker Institute for Cancer Immunotherapy and its venerable East Coast counterpart, the Cancer Research Institute, announced the formation of the Tumor Neoantigen Selection Alliance. This initiative, involving researchers from 30 universities, non-profit institutions and companies, aims to identify software that can best predict mutation-associated cancer antigens, also known as neoantigens, from patient tumor DNA. The hope is that solving the shortcomings of current *in silico* methods for identifying neoantigens will galvanize a new wave of personalized cancer immunotherapies. But, for now, computational prediction of neoantigens capable of eliciting efficacious antitumor responses in patients remains a hit-or-miss affair.

Cancer vaccines have traditionally targeted tumor-associated self-antigens—proteins that may be aberrantly expressed in cancer cells. More recently, however, attention has shifted to neoantigens. Targeting an individual's tumor-specific mutations is attractive because these peptides are new to the immune system and are not found in normal tissues. Compared with tumor-associated self-antigens, neoantigens elicit T-cell responses not subject to host central tolerance in the thymus and also produce fewer toxicities arising from autoimmune reactions to non-malignant cells.

The process of identifying candidate neoantigens in a patient starts with exon sequencing of a cancer biopsy and normal tissue to pinpoint missense mutations occurring in tumor-expressed proteins. Often, transcriptome data are also added to indicate antigen abundance.

The key question for neoepitope discovery is which mutated proteins are processed into 8- to 11-residue peptides by the proteasome, shuttled into the endoplasmic reticulum by the transporter associated with antigen processing (TAP) and loaded onto newly synthesized major histocompatibility complex class I (MHC-I) for recognition by CD8⁺ T cells.

Although some computational methods focus on predicting what happens during antigen processing (e.g., NetChop) and peptide transport (e.g., NetCTL), most efforts focus on modeling which peptides bind to the MHC-I molecule. Neural network–based methods, such as NetMHC, are used to predict antigen sequences that generate epitopes fitting the groove of a patient's MHC-I molecules. Other filters can be applied to deprioritize hypothetical proteins and gauge whether a mutated amino acid either is likely orientated facing out of the MHC (toward the T-cell receptor) or reduces the affinity of the epitope for the MHC-I molecule itself.

All kinds of confounding factors mean that these predictions can go awry. Sequencing already introduces amplification biases and technical errors in the reads used as starting material for peptides. Modeling epitope processing and presentation also must take into account the fact that humans have ~5,000 alleles encoding MHC-I molecules, with an individual patient expressing as many as six of them, all with different epitope affinities. Methods such as NetMHC typically require 50–100 experimentally determined peptide-binding measurements

for a particular allele to build a model with sufficient accuracy. But as many MHC alleles lack such data, 'pan-specific' methods—capable of predicting binders based on whether MHC alleles with similar contact environments have similar binding specificities—have increasingly come to the fore.

Today, a raft of software tools for predicting MHC binders are now available (http://cancerimmunity.org/resources/webtools/). But each of these packages has its own idiosyncrasies, strengths and weaknesses. What's more, it has proven difficult to benchmark which tools and combinations of tools work best for particular contexts.

The Tumor Neoantigen Selection Alliance seeks to address these challenges. In a partnership with the open science non-profit Sage Bionetworks, research groups participating in the alliance will receive genetic sequences from both normal and cancerous tissues. Using their own algorithms, each group will output a set of predicted neoantigens that will then be validated to assess which predicted peptides are presented by the MHC-I molecule of interest and recognized by T cells.

The availability of Sage's Synapse platform to facilitate sharing of cancer neoepitope binding data is an important step forward. Greater access to experimental measurements of peptide binders of MHC molecules will improve algorithm accuracy. At the moment, tens of thousands of peptides, eluted from the MHCs and identified by mass spectrometry (MS) lie dormant in the depths of supplementary data in papers and laboratory computers.

But releasing the data will not be enough.

Most neoepitope studies identify thousands of somatic mutations and predict a few hundred or so MHC binders. But typically the vast majority of these neoepitopes fail to turn up in tumors when tested for using MS; worse still, only a handful are found to elicit T-cell responses. Why is this?

One reason is that methods for assessing neoepitope immunogenicity are less than perfect, relying on MHC-peptide tetramer staining of tumor-infiltrating lymphocytes (TILs). Also, the small numbers of TILs that can be isolated and the lack of TILs in some 'cold' tumors mean lymphocytes must often be expanded *ex vivo*, which can alter T-cell specificities.

Another problem is that a critical arm of cellular immunity is often overlooked: most *in silico* tools focus on MHC-I binders. In contrast, algorithms for predicting MHC-II binders (important for CD4⁺ T-cell antitumor responses) have lagged because motifs are more promiscuous, epitopes are longer in length and more variable, and more data are required to underpin models.

The truth is then that current neoepitope prediction algorithms return a vast number of candidates, of which only a tiny handful are ever found to trigger *bona fide* antitumor responses in patients. Despite the profundity of cancer cell mutations, immunogenic neoantigens are the exception rather than the rule. This means there is a great deal more research to do before neoepitope prediction and validation becomes routine and personalized immunotherapy a clinical reality.