Immuno-Oncology Therapies and Precision Medicine: Personal Tumor-Specific Neoantigen Prediction by Machine Learning

Yi-Hsiang Hsu, MD, SCD
Sep 16, 2017
yihsianghsu@hsl.Harvard.edu

HSL GeneticEpi Center, Harvard Medical Sch
Program for Quantitative Genomics, Harvard School of Public Health
BROAD Institute of MIT and Harvard
NHLBI Framingham Heart Study Investigator
Outline

- Background on Cancer Immune Therapies
- Problems
- Ongoing Works on Neoantigen Predictions by Deep Learning and Validations
Challenges in Treating Cancers

- Every cancer is different (inter-cancer type heterogeneity)
- Every cancer patient is different (inter-tumor heterogeneity)
- Every tumor cells in a tumor site can be different (intra-tumor heterogeneity)

*Genetic and phenotypic variation

*Intertumour heterogeneity

*Intratumour heterogeneity

*Clonal heterogeneity

*Intercellular genetic and non-genetic heterogeneity

Carter had melanoma that had spread to his liver and brain
Pembrolizumab, PD-1 (programmed cell-death 1) checkpoint inhibitor
Immunotherapies

Coley’s Toxin
Serratia
Streptococcus

Immune checkpoints
The Cancer–Immunity Cycle

Major Histocompatibility Complex Alleles (MHC)
Normaly, the immune system should recognize tumor cells and distinguish them from their normal counterparts.

In cancer patients, tumor cells escape from immune system surveillance via immune checkpoints (inhibitory pathways to inactivate T-cells).

Under normal physiological conditions, immune checkpoints are crucial for the maintenance of self-tolerance (that is, the prevention of autoimmunity).
Cancer Immunotherapies

- Immune Checkpoint Modulators
  Block immune checkpoint proteins, such as CTLA4, PDL-1, PDL2 that inactivate T-cells

- Immune System Modulators
  Uses proteins that normally help regulate, or modulate, immune system activity to enhance the body’s immune response against cancer. These proteins include cytokines and certain growth factors. Two types of cytokines are used to treat patients with cancer: interleukins and interferons

- Therapeutic antibodies:
  Antibodies made in the laboratory that are designed to cause the destruction of cancer cells
  Antibody–drug conjugates (ADCs): chemically linking antibodies to a toxic substance. The antibody portion of the ADC allows it to bind to a target molecule that is expressed on the surface of cancer cells. The toxic substance can be a poison, a small-molecule drug or a radioactive compound. Once an ADC binds to a cancer cell, it is taken up by the cell and the toxic substance kills the cell
Cancer Immunotherapies

- **Immune Cell Therapy: CAR T-cells:** Patient's T cells are collected from the blood and genetically modified to express the chimeric antigen receptor, or CAR. The modified cells are grown in the laboratory to produce large populations of the cells, which are then infused into the patient.

- **Cancer Vaccines:** These vaccines are usually made from a patient’s own tumor cells or from substances produced by tumor cells. They are designed to treat cancers that have already developed by strengthening the body’s natural defenses against the cancer.

- **Personalized Tumor-Specific Neoantigen-Targeted Vaccines or T-cell-based Therapy**
  - Tumor somatic mutation-derived antigens (neoantigens)
  - Neoantigens are specifically expressed in the tumor, but not in normal cells
  - More “foreign” to the immune system and stronger immunogenicity than tumor-associated “self” antigens
  - Less likely to induce tolerance
  - Nearly impossible to induce normal tissue toxicity
  - **Adaptable, durable and synergistic**
Tumor-Cell Specific Neoantigens
What Make A Somatic Mutation A Neoantigen?

- A consequence of tumor cell mutations and express as abnormal proteins
- These proteins partially degraded by the normal cellular re-cycling machinery, creating short 8 to 12 amino acid peptides
Estimate of the Neoantigen Repertoire

- Estimated based on the frequencies of somatic mutations

Schumacher and Schreiber, Science, 2015
Estimate of the Neoantigen Repertoire

- Estimated based on the frequencies of somatic mutations

- A large fraction of the neoantigens in human tumors is not shared between patients at meaningful frequencies:

  Requires personalized immunotherapies

Schumacher and Schreiber, Science, 2015
A Successful Case Study:
A 50-year-old woman with metastatic colorectal cancer (ClinicalTrials.gov number, NCT01174121)

Adoptive transfer of ex vivo expanded tumor-infiltrating lymphocytes containing T cells (CD8+) targeting personalized cancer neoepitopes to mediate regression of metastatic solid cancers

KRAS G12D: glycine (G) to aspartic acid (D):
The most frequent KRAS mutant in gastrointestinal cancers
45% of pancreatic cancers and 13% of colorectal cancers

A single infusion of 1.48×10^{11} tumor-infiltrating lymphocytes:
~75% CD8+ T cells (1.11×10^{11} cells) reactive to mutant KRAS G12D

Tran, et al., NEJM 2016
Findings

Tran, et al., NEJM 2016
Somatic Mutations in Cancer Patients

Schumacher and Schreiber, Science, 2015
Accurate prediction of neoantigens:
Neoantigens are immunogenic epitopes

→ High binding affinity of mutated peptides to MHC molecules
  (High peptide-MHC binding density on tumor cell/APC surface)

→ CD8+ or CD4+ T-cell receptors (TCRs) recognize of mutated peptide-MHC binding complexes
Ongoing Projects
Deep Learning


- It became popular recently:
  - New learning algorithms
  - Advances in hardware like GPU
  - Large and complex datasets

- Different forms such as DBN, CNN, RNN...
Convolutional Neuron Network (CNN) Model for Mutated Peptide-MHC Binding Prediction

- Input data
  - Interaction between HLA-A0201 N-terminal (180 aa) and peptide (9 aa)
  - AA interactions are from AAindex, TANS760101; epitopes from IEDB

- Model structure
  - Two convolution and pooling layers, as in the figure to the left
  - One fully connected layer (Fc)
  - Drop out rate of Fc layer: 0.4

- Prediction
  - Transformation of IC$_{50}$, i.e., $1 - \log(\text{IC}_{50})/\log(50000)$
Comparison of Individual Models to the Pan-Model

Pearson Correlation

Accuracy

Individual  Pan-model
A platform to compare and validate neo-epitope prediction pipelines

Includes validation of predicted epitopes using both binding and functional assays
The TESLA Project: Contributors

- Co-Hosts: CRI and Sage
- 22 Academia/Non-Profits
- 12 Pharma/Biotech

Institutions:
- MD Anderson Cancer Center
- Memorial Sloan Kettering Cancer Center
- Stanford Medicine
- UCLA
- UCSF
- Penn Medicine
- Caltech
- La Jolla Institute
- National Cancer Centre Singapore
- New York Genome Center
- NIBIT
- Roswell Park
- UConn Health
- VA Medical Center, Topeka
- ViaCyte
- Advaxis
- Agensys
- Amgen
- Biontech
- Biotechs
- Bristol-Myers Squibb
- GlaxoSmithKline
- Genentech
- ISA Therapeutics
- MedImmune
- Medgenome
- Neovii
- Personalis
- University of California, Santa Cruz
The TESLA Project: Targeted Tumors

- Three tumor types with high load mutations
  - Melanoma – treated and untreated (UCLA)*
  - NSCLC – untreated (MSKCC)
  - CRC – untreated

- Melanoma: first tumor samples to be analyzed
  - 4 cases
    - DNAseq (tumor and normal)
    - RNAseq
  - Clinical HLA

- Other prioritized tumor types
  - Ovarian
  - Bladder
  - Triple negative breast
The TESLA Project: Validation

- Validation of epitopes
  - Ultimate validation via vaccination in humans
  - Initial validation via
    - Peptide binding – Key feature in algorithms and likely major driver for recognition
    - Epitope recognition by TILs and/or PBMC – Ultimate read-out for T cell recognition. Will test via both direct binding assessments as well as functional readouts

- Technical approaches
  - HLA binding (A. Sette, LIAI)
  - Tetramer binding (R. Schreiber, WUSTL)
  - Tetramer nanoparticle isolation (J. Heath, Caltech)
  - Cytokine production (N. Bhardwaj, Mt. Sinai)
  - Other – TBD
Summary and Future Work

- The benefits of training all HLA-A types together can be seen in two aspects
  - For rare alleles with little or no training data, this model provides a way to do the prediction
  - Even for common alleles like A0202, the prediction accuracy improved due to its AA sequencing similarity with other HLA-A type, such as the HLA A0201 allele

- CNN outperformed other simple version of NN, such as ANN, DBN and RNN

- Ongoing works:
  - Add additional 48 differential structural and biochemical information (matrices) into the model
  - Add additional HLA types (HLA-B, HLA-C)
  - Extend to allow mutated peptides ranged 9 to 15 aa
  - Re-train models with validation data
  - Predict binding on Peptide-MHC complex and T-cell receptors (TCRs)

yihsianghsu@hsl.Harvard.edu