Immunogenicity Analysis of Somatic Mutations

- 1. **translated all mutations in exomes so strings of 17 amino acids were generated for the predicted wild type and mutant, with the amino acid resulting from the mutation situated centrally.**
- 2. **evaluate MHC Class I binding, wild type and mutant nonamers containing the tetrapeptides common to the complete responders.**
  - 1. NetMHC v3.4
  - 2. RANKPEP
- 3. **assessed for similarity between nonamers that were predicted to be presented by patient-specific MHC Class I.**
  - 1. Weblogo
- 4. **evaluated nonamers for putative binding to the T cell receptor**
  - 1. IEDB immunogenicity predictor
  - 2. CTLPred
- 5. **evaluate homology(conserved tetrapeptides) to known pathogens' antigens**
  tetrapeptides assessed as substrings of immunogens in the database for a positive T cell response in Homo sapiens host. Tetrapeptides were compared to known antigens using the functions "linear epitope," "substring",Homo sapiens host and positive assay result.
  - 1. Immune Epitope Database
- 6. **the neoantigen landscape/signature analysis**
  - 1. either the tetrapeptide occurred 3 or more times in discovery set benefiters and not in non-benefiters
  - 2. or 2 times with resemblance to IEDB via the substring method described above
  - 3. Of note, all tetrapeptides were located within a 9 amino acid stretch predicted to be presented by each patient's HLA. (值得注意的是，所有的四肽都位于每个患者HLA预测的9个氨基酸的范围内。)
  - 4. Hierarchical clustering of the tetrapeptide motifs is used to help identify the most frequent motifs in the discovery set responders.
  - 5. Regression is used to order the tetrapeptides to confirm the ones that tracked most with clinical benefit. (***This reduced the candidate tetrapeptides from 4,325 to 216. And hese 216 peptides represent a locked set from the discovery set.***)