High-throughput neoantigen discovery pipeline using noncanonical peptides

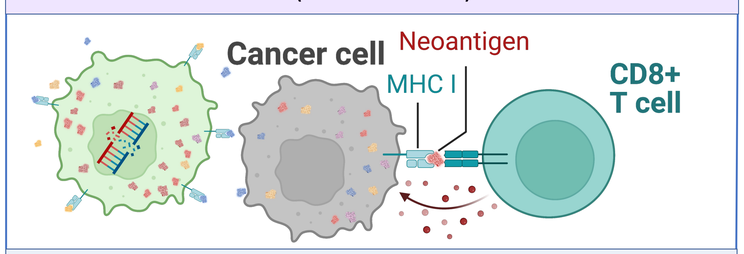
Siddharth Mahesh[1], Ali Mobedi[1], Chenghao Zhu[2], Lydia Y. Liu[2], Paul C. Boutros[2]

1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA
2 Department of Human Genetics, David Geffen School of Medicine, UCLA

Abstract

The study of tumor neoantigens is a new direction in cancer immunotherapy that relies on the body's immune system to eliminate cancerous cells[2]. Neoantigens are peptides derived from genomic and transcriptomic variations that are recognized by the major histocompatibility complex (MHC). These tumor-specific peptides can then be presented on the cell's surface for identification by T-cell receptors. Recent studies demonstrate the low accuracy of current neoantigen discovery pipelines, revealing that less than 5% of predicted bound peptides are truly present on cell surfaces[3].

We propose a well-optimized high-throughput pipeline that uses a non-canonical peptide caller (via moPepGen), HLA-genotyping (via OptiType), and MHC class I binding affinity predictor (via MHCFlurry) in neoantigen discovery. The pipeline was tested on sequencing data obtained from 14 metastatic head and neck squamous cell carcinoma tumors (HNSC tumors).



Methods

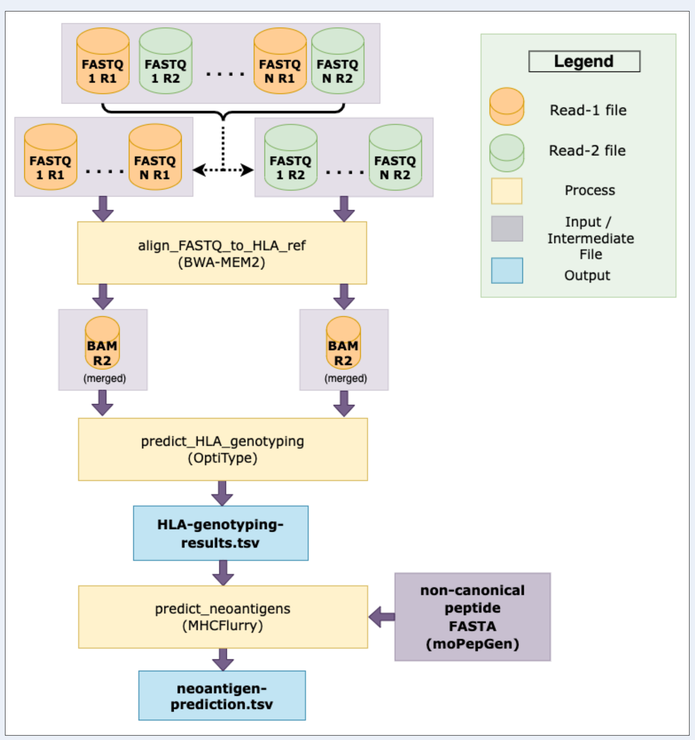


Figure 1: Workflow of the pipeline, given a single patient, starting with whole genome (WGS) or whole exome (WXS) sequencing data.

Results

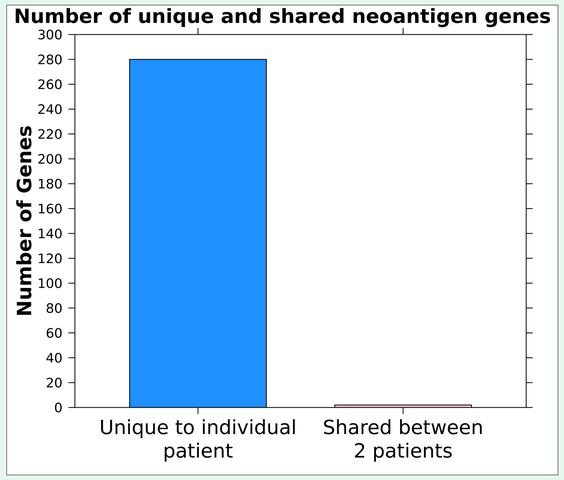


Figure 2: Most neoantigen genes are patient-specific. Out of the total 282 genes contributing to neoantigens, 280 genes were unique to individual patients while only 2 genes were shared between two patients.

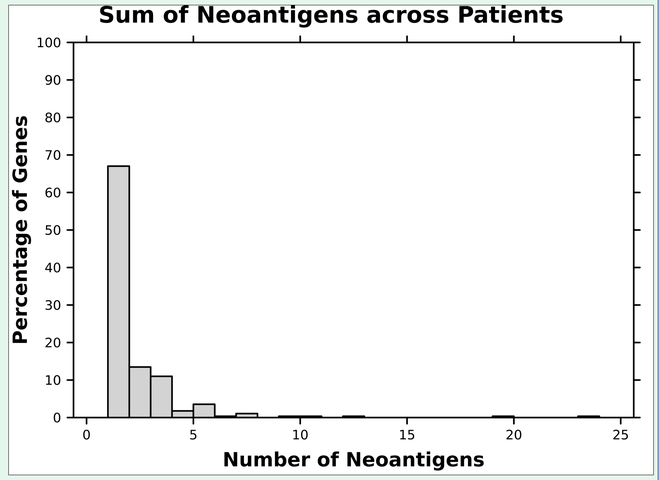


Figure 3: Most neoantigen genes contribute to less than 5 neoantigens. The sum of neoantigen numbers per gene is calculated for all 14 patients.

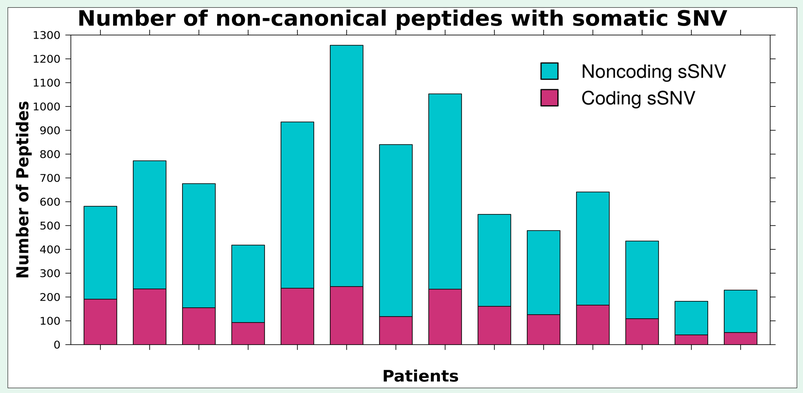


Figure 4: Most non-canonical peptides are derived from noncoding somatic SNVs. The moPepGen pipeline indicates whether the detected peptide belongs to a coding or noncoding transcript.

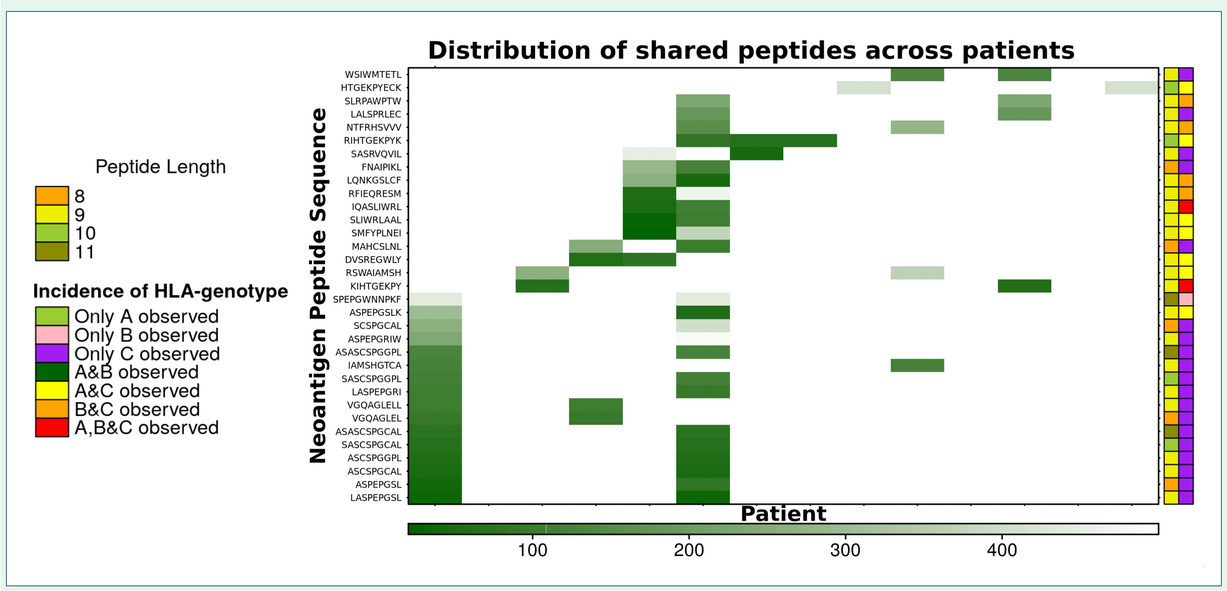


Figure 5: Analyzing the distribution of peptides observed across multiple patients. Peptide length and incidence of various HLA-genotypes (shown on right) are covariates for each neoantigen peptide. Binding affinity between the given neoantigen peptide and MHC-1 protein complex is denoted by cell color, where darker colors indicate higher binding affinity.

Conclusions

We developed a high-throughput pipeline that successfully ran on sequencing data obtained from 14 HNSC patients. Within this dataset, we found that:

- Most neoantigen genes are unique to individual patients **(Fig 2).**
- The number of neoantigens per mutated gene shows a skewed distribution, where a few of genes yield many potential neoantigens **(Fig 3).**
- Most non-canonical peptides are derived from noncoding somatic SNVs **(Fig 4).**
- Only 0.38% (33/8792) of all neoantigens discovered were shared between multiple patients (**Fig 5**).
  - All these shared peptides had a length between 8-11 amino acids **(Fig 5).**

Overall, by leveraging proteogenomic data, we hope that our pipeline can ultimately be a powerful tool for accurate neoantigen prediction in the development of cancer vaccines, improvement of patient outcomes, and disease prognosis.

Future Work

In the future, we can:

- verify if the predicted neoantigens are truly bound to the MHC-1 protein complex (using LC-MS).
- compare predicted neoantigens with other affinity predicting pipelines.
- compare samples of multiple tumor types to investigate whether the previously mentioned trends are tumor-specific.

References

1. Cells were created using BioRender.com
2. Bai, Ri-Lan; Chen, Nai-Fei; Li, Ling-Yu; Cui, Jiu-Wei. A brand new era of cancer immunotherapy: breakthroughs and challenges. Chinese Medical Journal: June 05, 2021 - Volume 134 - Issue 11 - p 1267-1275 doi: 10.1097/CM9.0000000000014904
3. Zhang Z, Lu M, Qin Y, Gao W, Tao L, Su W and Zhong J (2021) Neoantigen: A New Breakthrough in Tumor Immunotherapy. *Front. Immunol.* 12:672356. doi: 10.3389/fimmu.2021.672356