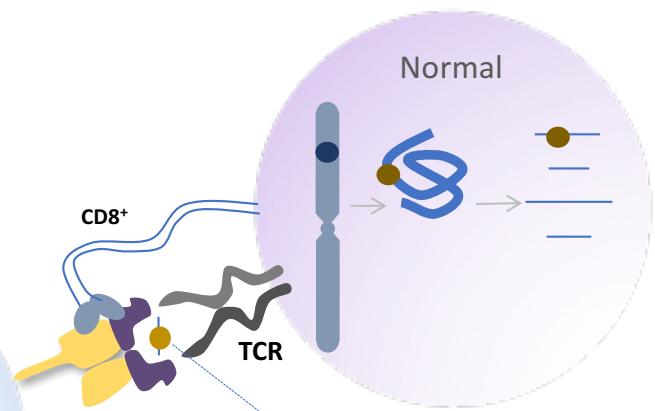


# ML based Prediction of Immunogenic Neoantigens for immunotherapy

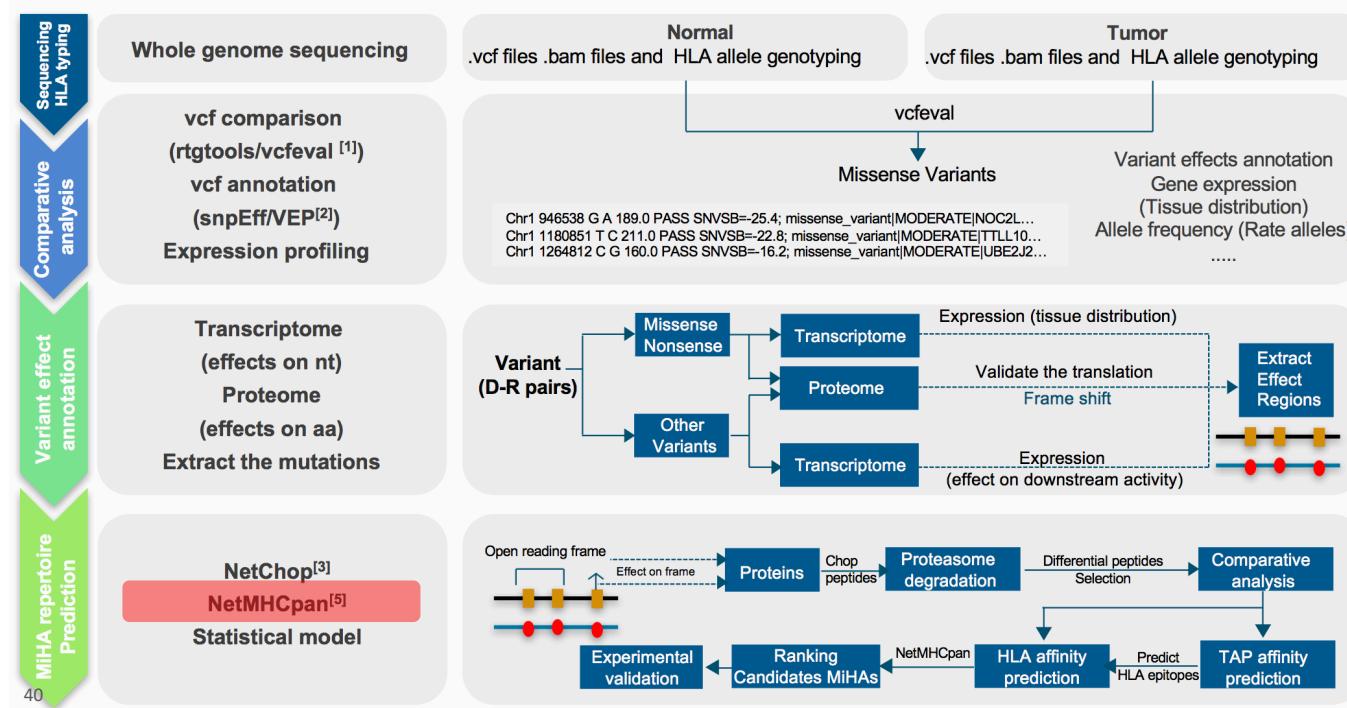
Team: bearlyunderstanding

Guy Wilson  
Joshua Price  
Alex Yao  
Forrest Li  
Lei Hung  
Wei Wang

SVAI Hackathon, SF  
6-25-2017



GitHub: <https://github.com/SVAI/bearlyunderstanding>

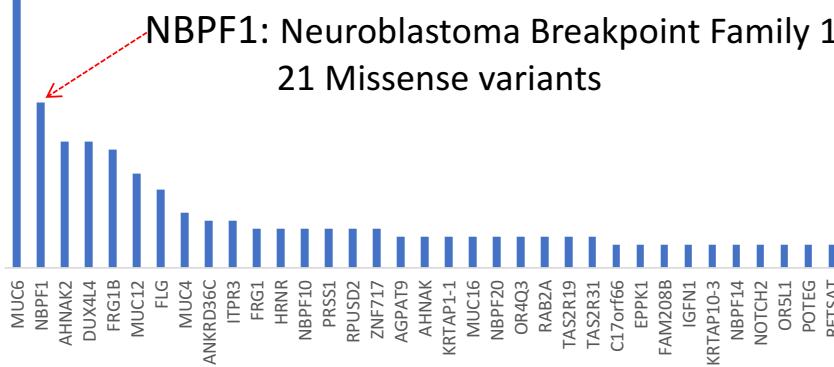


Available: [github.com/wwang-nmdp/MiHAIP](https://github.com/wwang-nmdp/MiHAIP)

Lack of HLA typing results?

HLA-A\*02:01 (Common allele)

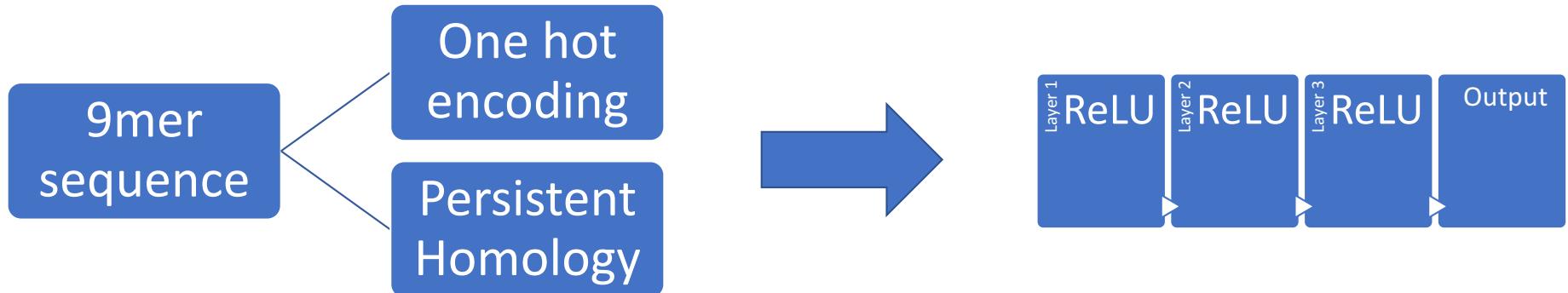
Prediction of HLA-restricted peptide



	Nucleotide	Peptide	IC50	%Rank
Tumor	A	YPQ <b>F</b> SDTLEL	123.1	1.2
Normal	C	YPQ <b>C</b> SDTLEL	1394.1	4.5

# Features, Network Design, Performance

- Feedforward ANN with 3 hidden layers (20, 20, 10)
  - Adagrad with light L2 regularization (0.001) and  $\alpha = 0.01$
- Results:
  - Avg MSE: 3148181632.000 (one hot encoding only)
  - Avg MSE: 3110076083.200 (persistent homology)



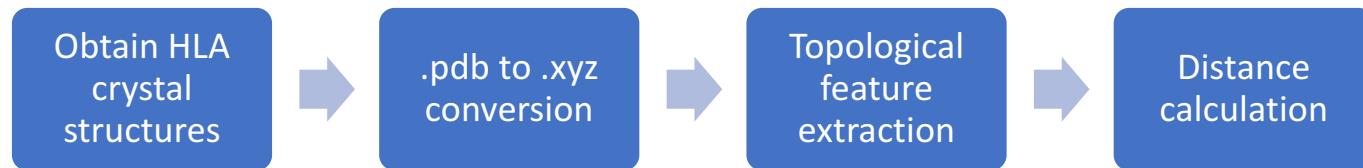
# Topological Feature Extraction

## Problem

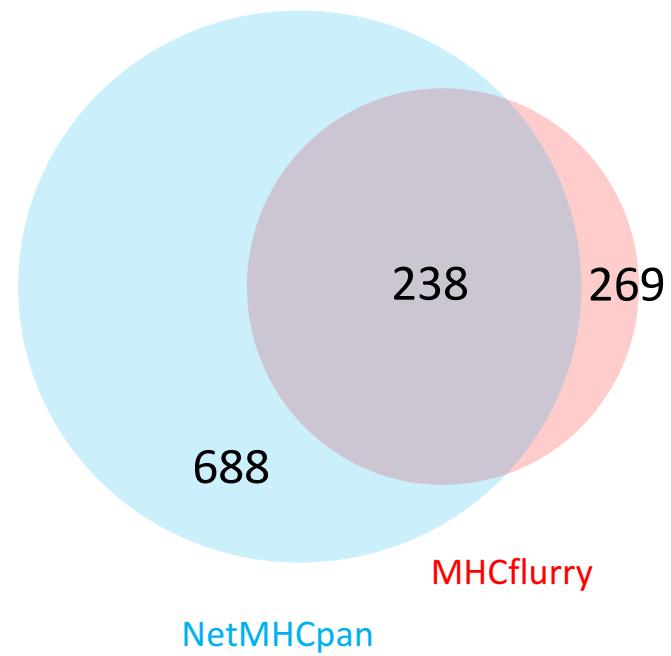
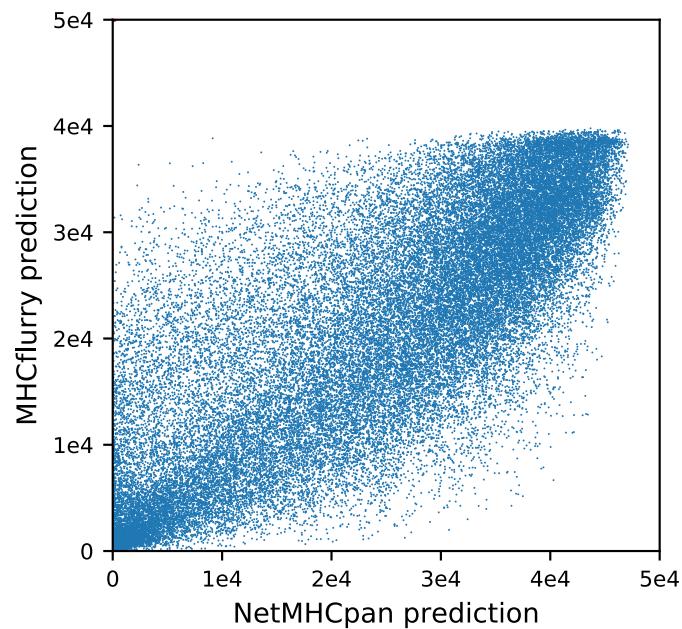
- Datasets have many ~lovely~ features
  - What is proper scale?

## Solution

- Consider all of them
  - Data as approximation of underlying manifold



## Comparison of Existing Methods: NetMHCpan and MHCflurry



Overlap of tight-binding peptides predicted  
by the two algorithms

## Conclusions:

- The developed tool can predict HLA class I restricted Neoantigens.
- Existing methods produce consistent predictions for strong-binders.
- With many grains of salt, topological methods may enable better performance in ML based epitope predictions.