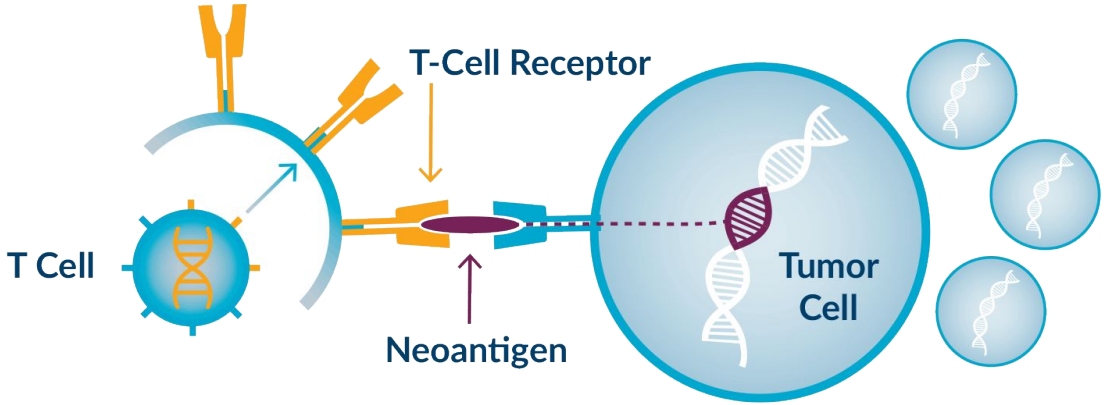
Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification

Bulik-Sullivan et al.

Presented by Wesley Wei Qian

Introduction - neoantigents



T-Cell receptors
T-cell receptors or TCRs are molecules on the surface of cancer fighting T cells that have the ability to interrogate individual cancer cells and see beneath the cell membrane.

Neoantigens
Antigens are the unique molecules or proteins that help immune cells identify and fight cancer cells. Neoantigens are unique to each patient's tumor cells.

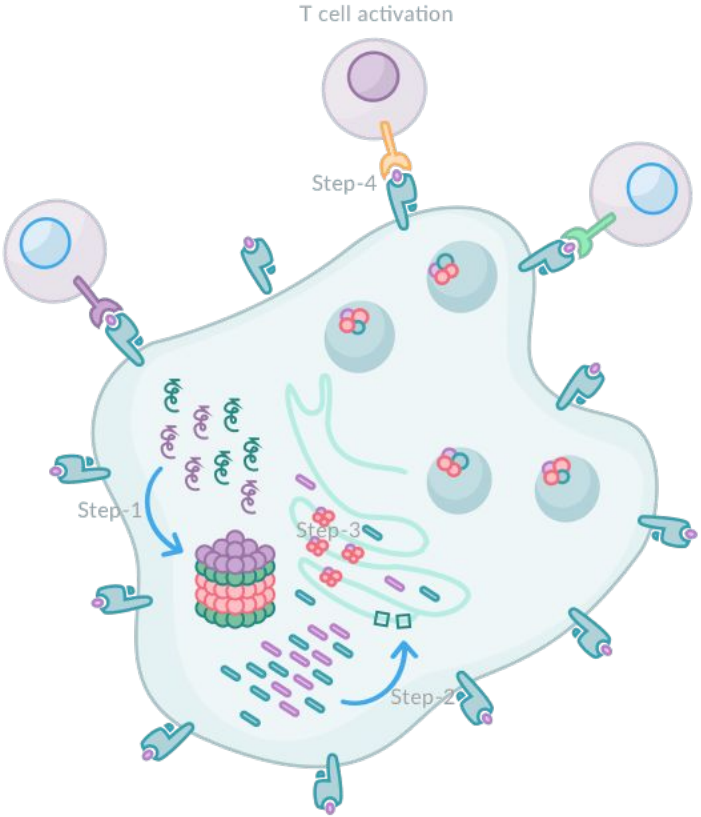
TCRs have the potential to be genetically modified so T cells can identify the neoantigen signature unique to one patient's cancer cells and then attack them.

https://ziopharm.com/t-cell-therapy/tcrs-for-solid-tumors

Introduction - the challenges

We cannot just look at all the mutations in the tumor cell

- Lots of them
- Many processes are involved before the neoantigents are presented in the cell surface
- Different HLA/MHC proteins



Steps in antigen presentation

https://research.medgenome.com/immuno-oncology-solutions/neo-epitope-prioritization-analysis/

Given a peptide, what's the probability for it to present in the surface as a neoantigent?

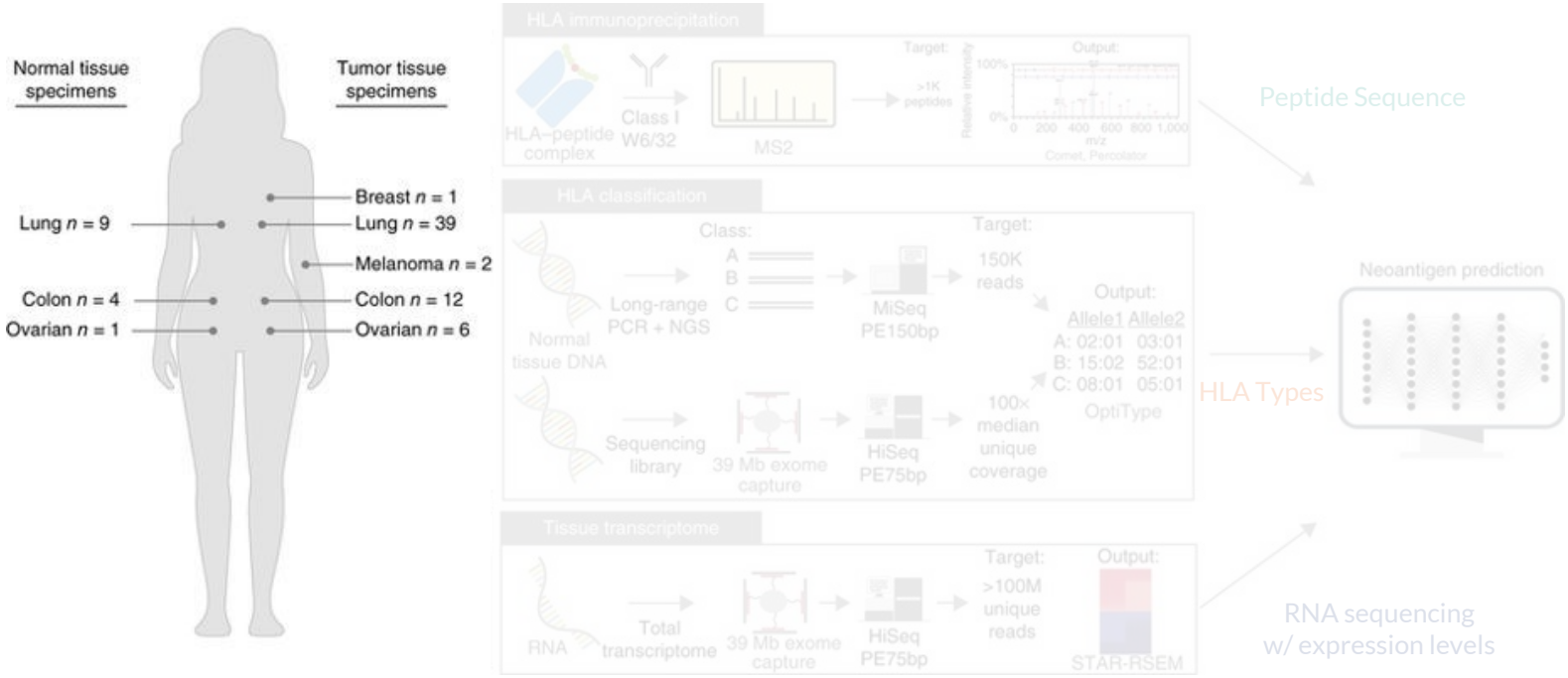
Introduction - previous approach

Predicting the peptide-HLA binding affinity

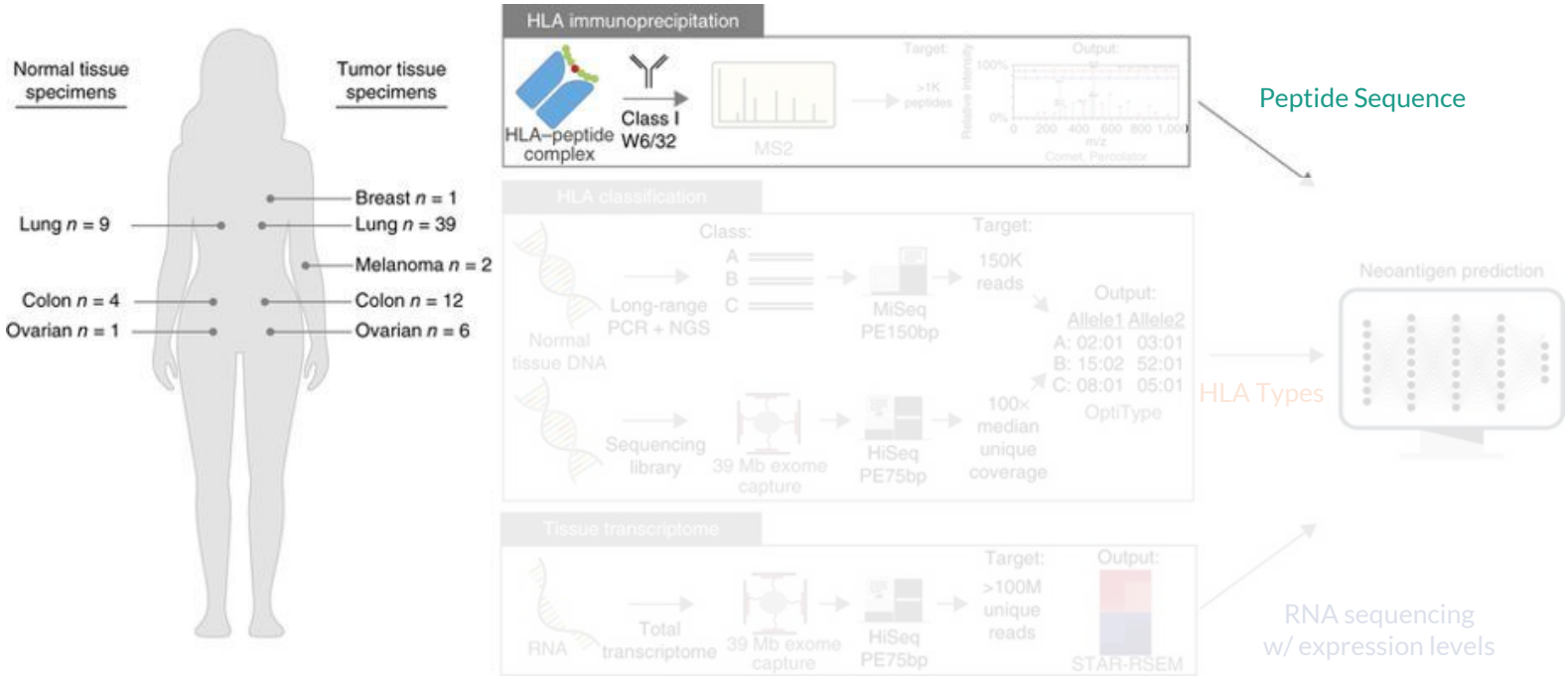
- State of the art performs very well in binding affinity prediction
- But lags behind in predicting actual presentation
- Peptide-HLC binding is only one step in the pathway

- Lots of information to consider in the process!

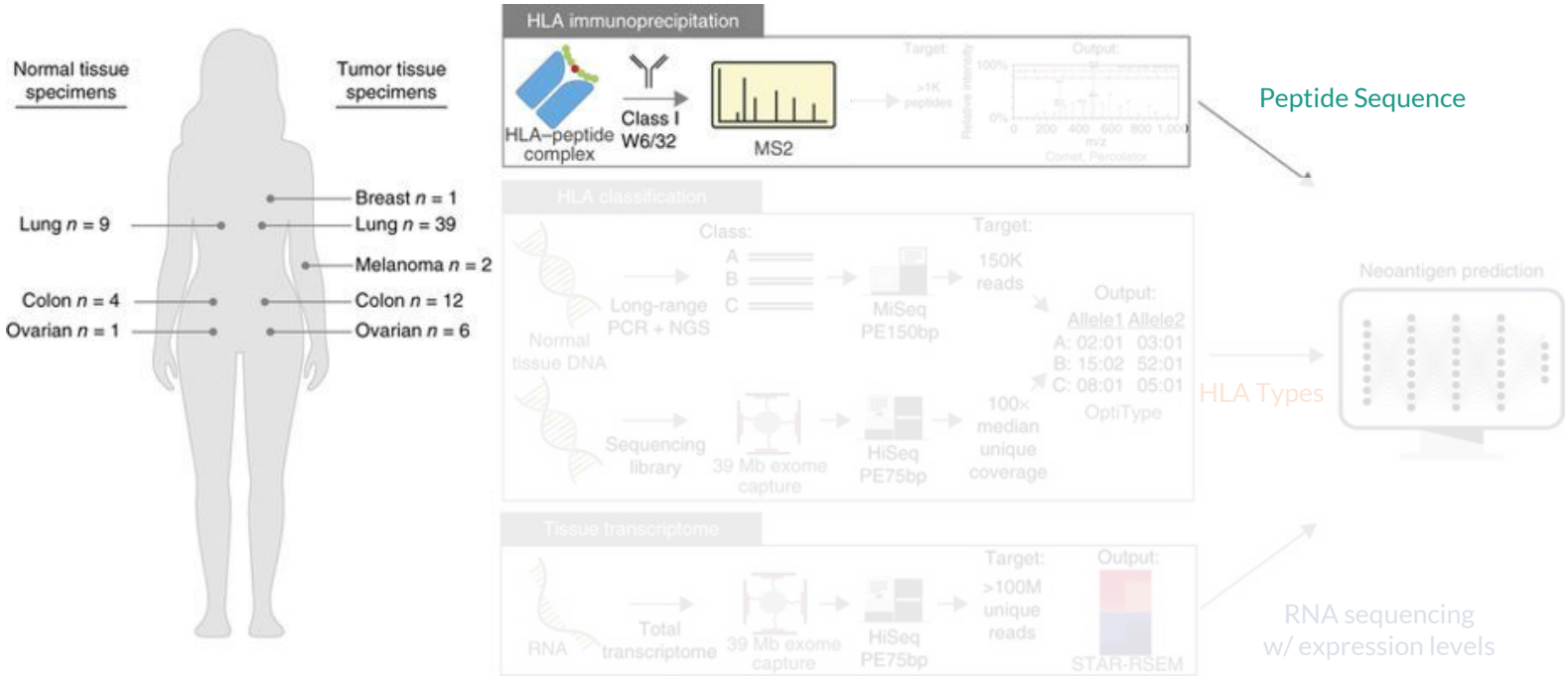
Methods - what data/input/features to consider



Methods - what data/input/features to consider



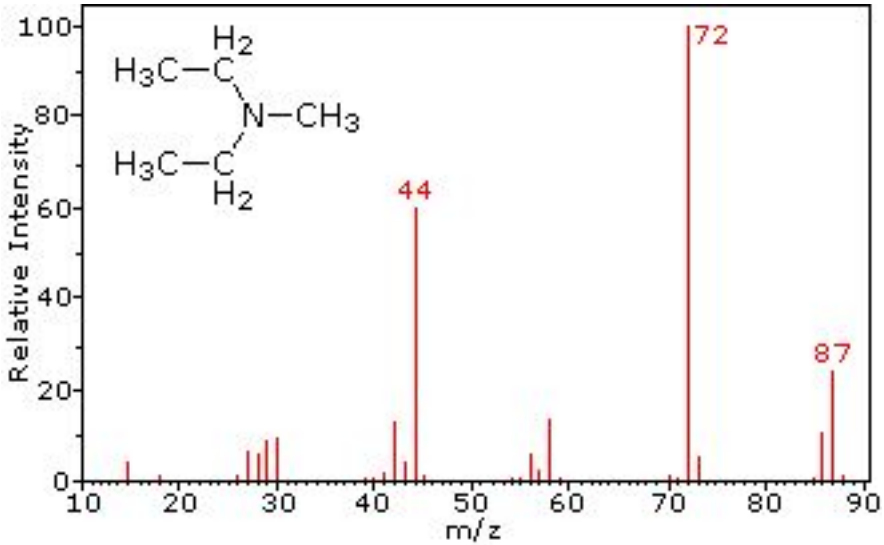
Methods - what data/input/features to consider



Methods - mass spectrometry

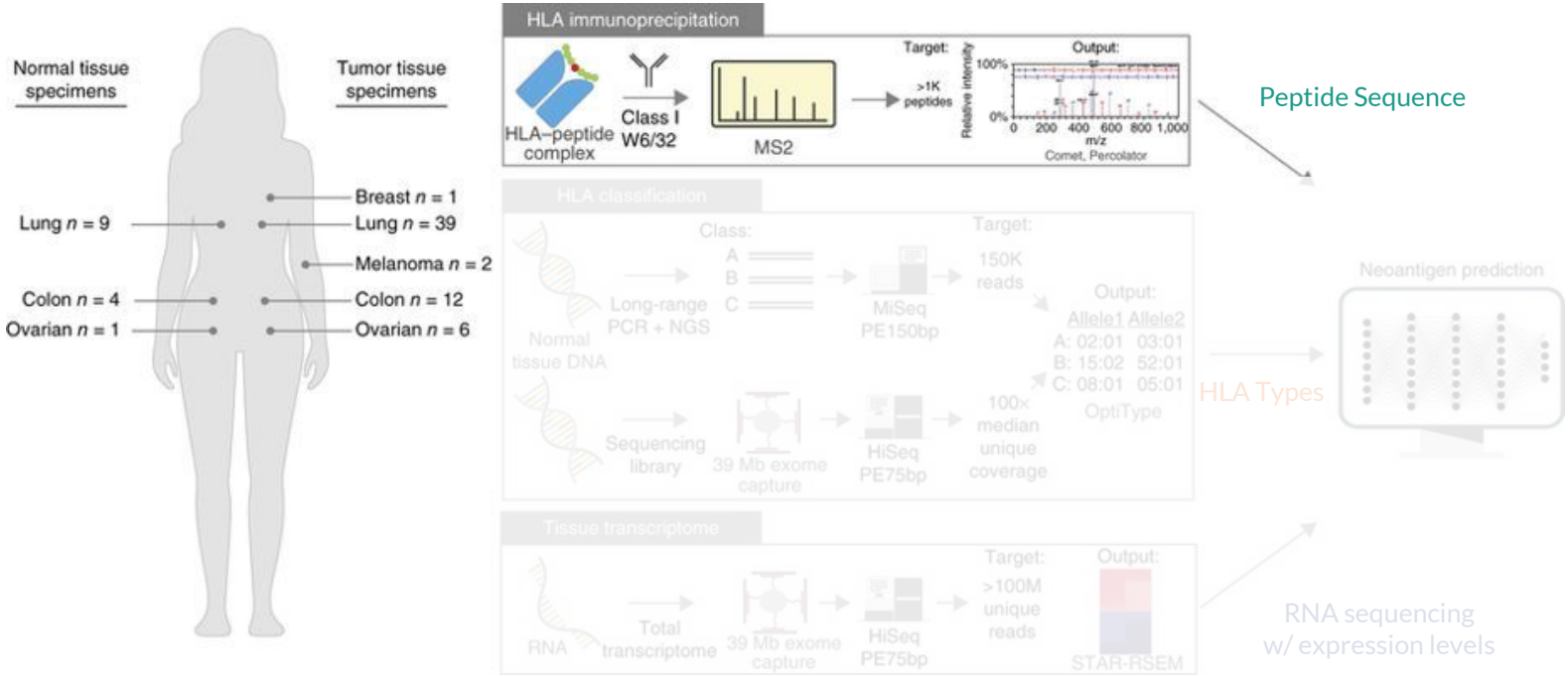
"Ionizing chemical species and sorting the ions based on their **mass-to-charge ratio**".

As a fingerprints to identify molecules/peptides.



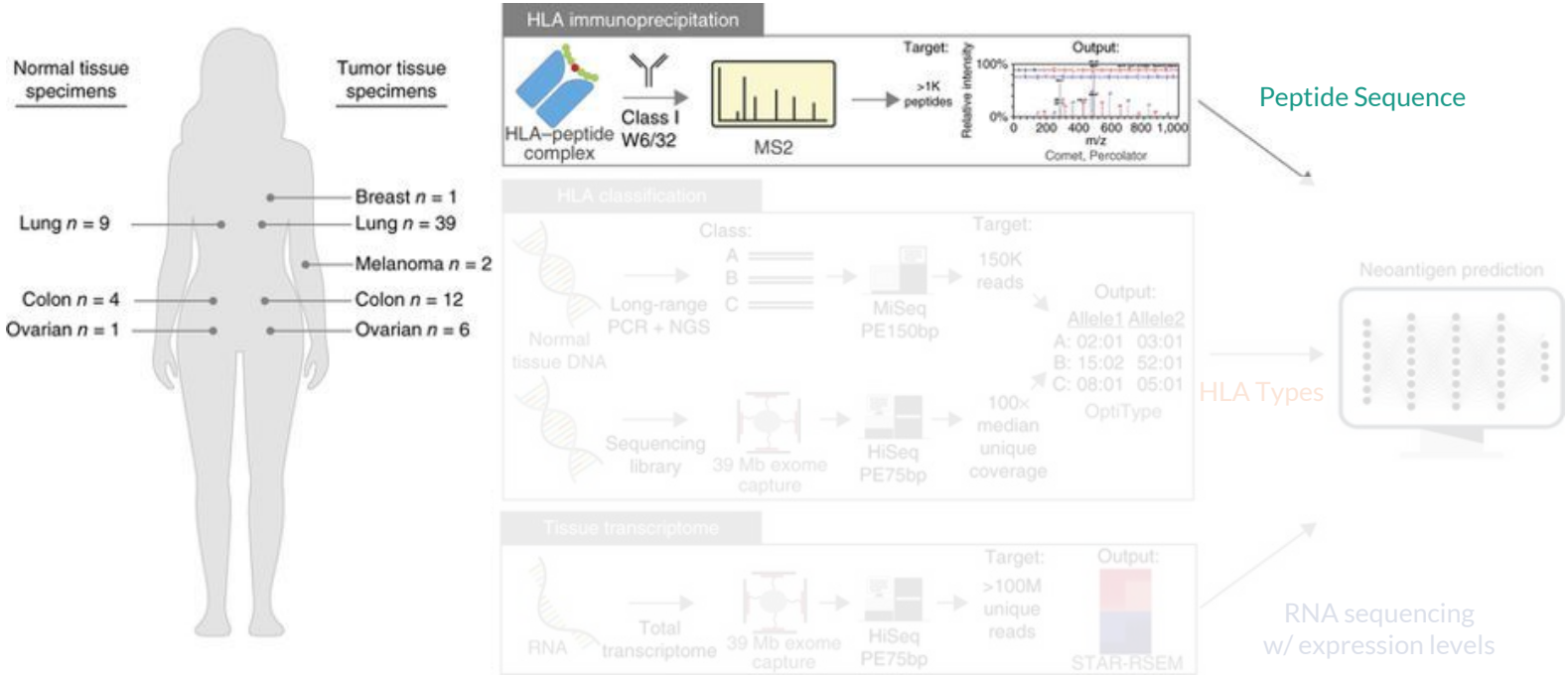
https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/massspec/masspec1.htm

Methods - what data/input/features to consider

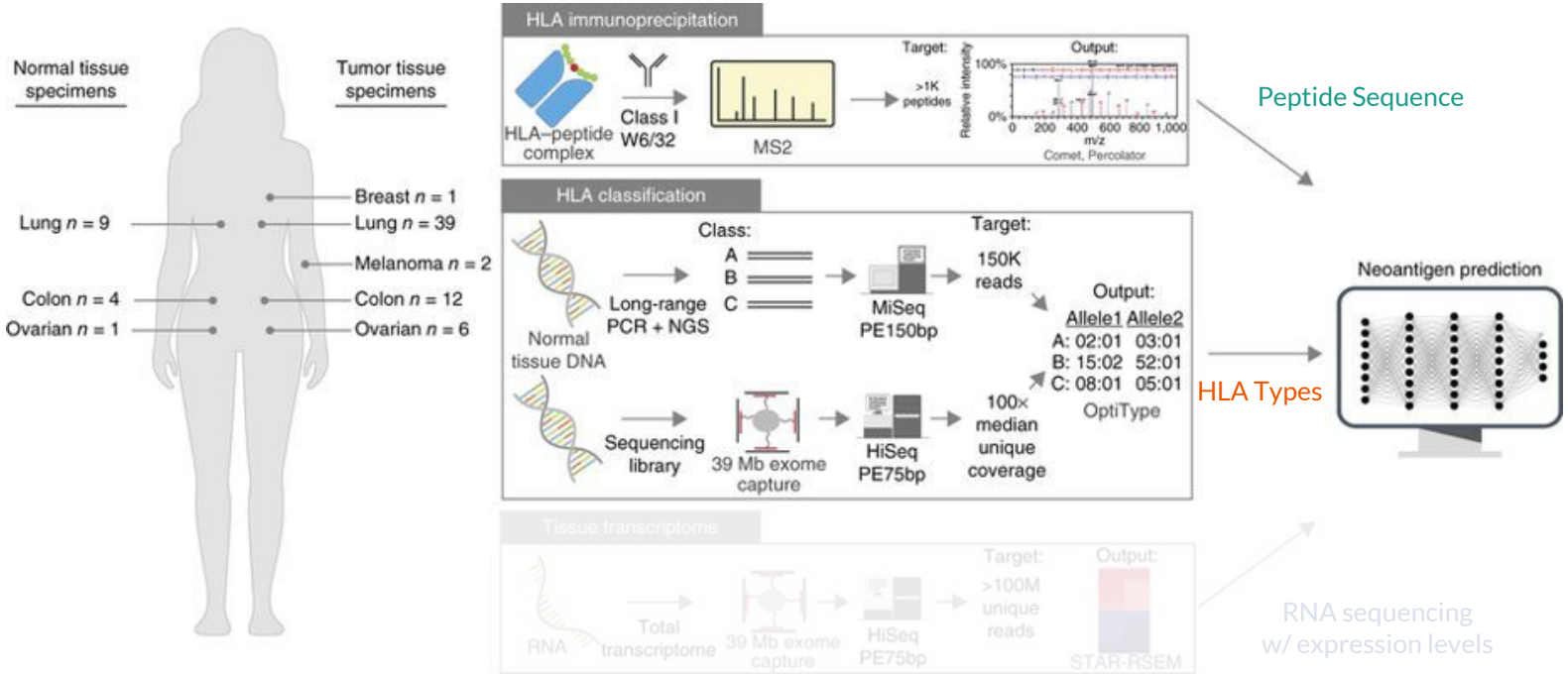


Methods - what data/input/features to consider

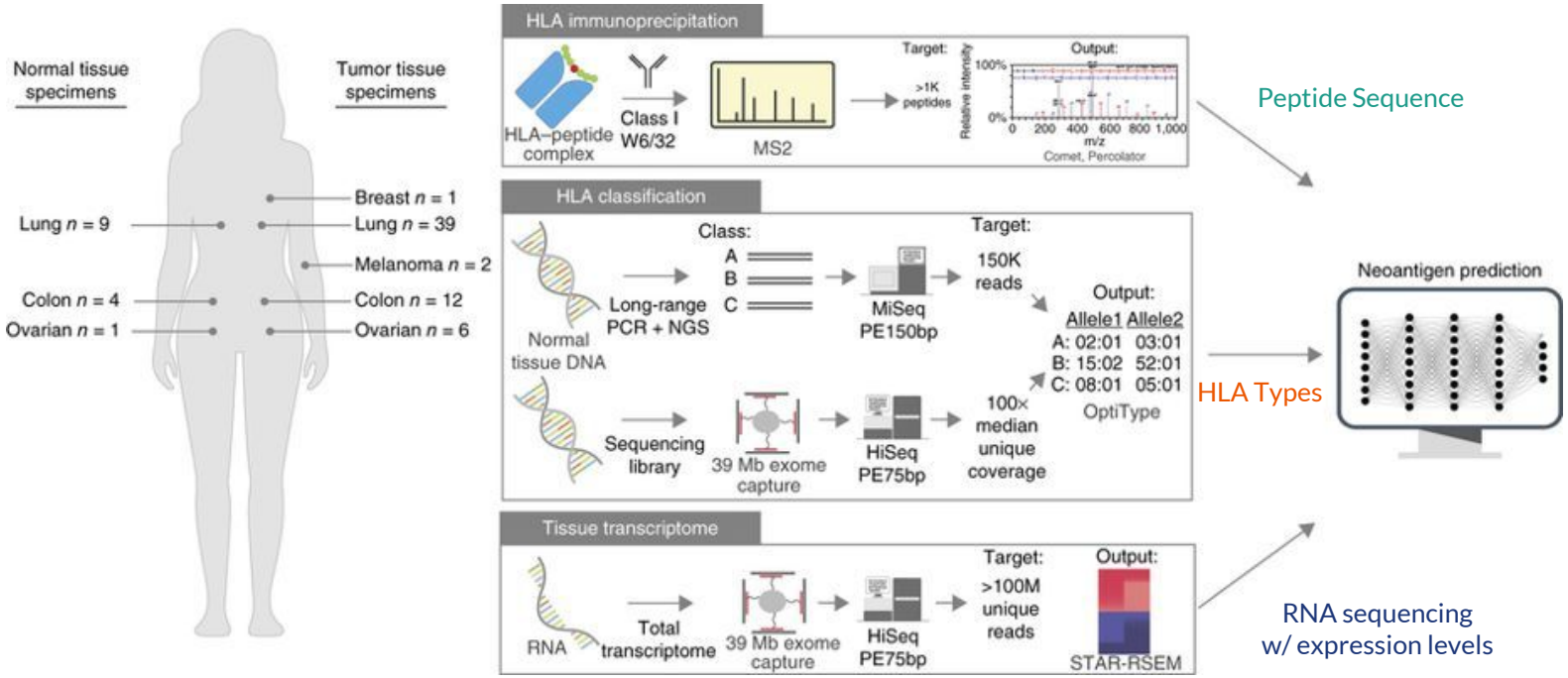
Positive Label! So what about Negative Label?



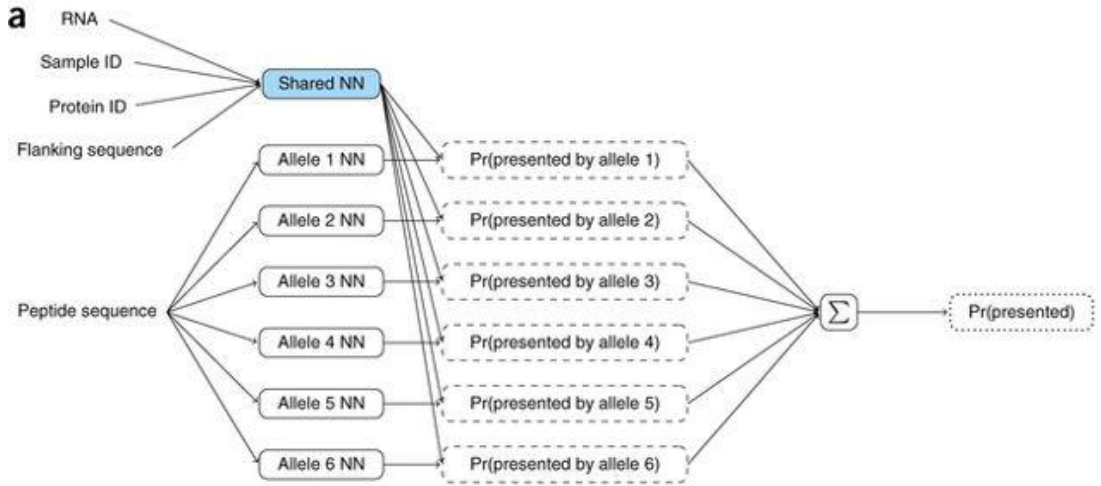
Methods - what data/input/features to consider



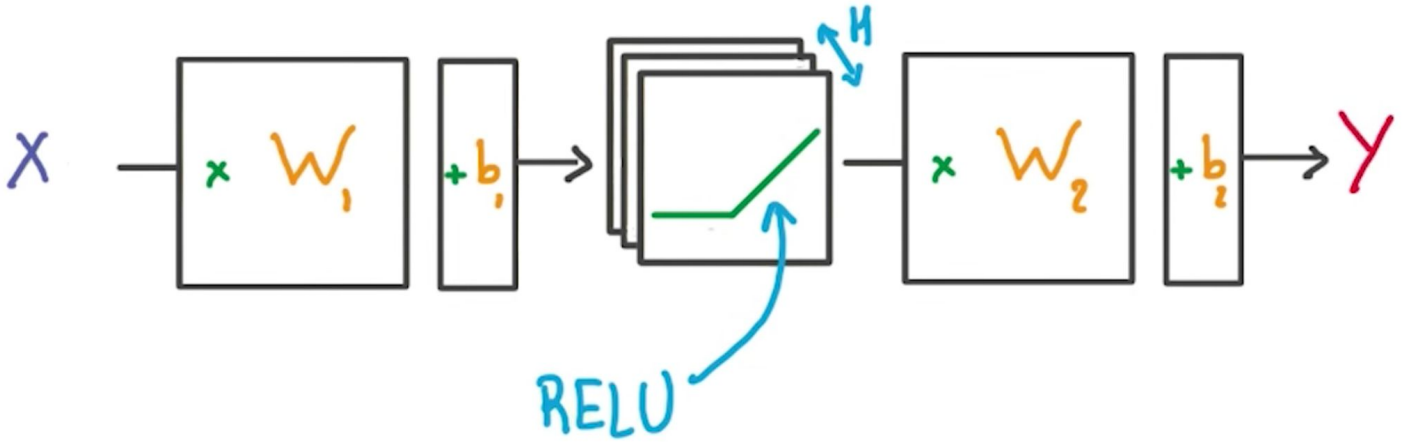
Methods - what data/input/features to consider



Methods - neural network model

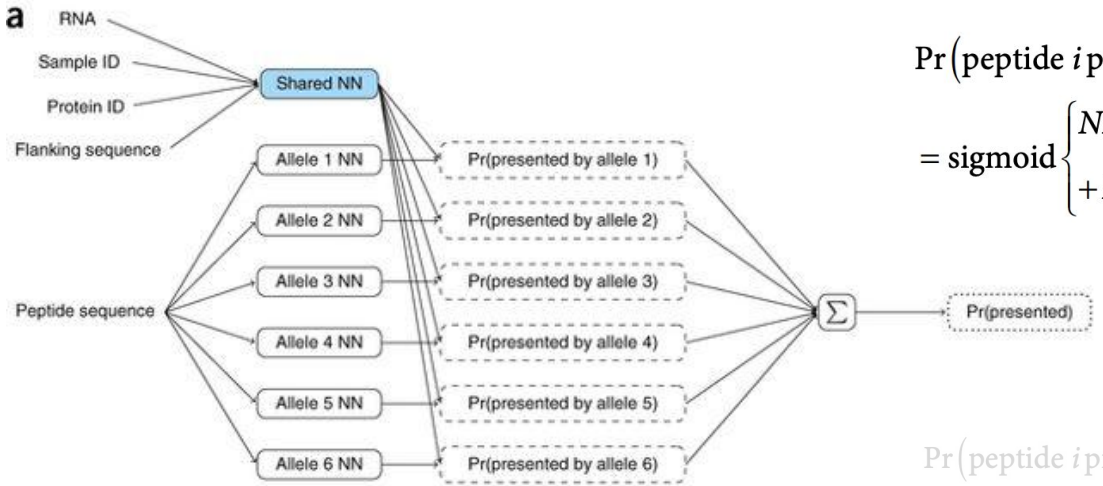


Methods - neural network is a function approximator easy to be trained



https://github.com/nehal96/Deep-Learning-ND-Exercises/blob/master/Deep%20Neural%20Networks/tf-deep-neural-networks.md

Methods - final prediction

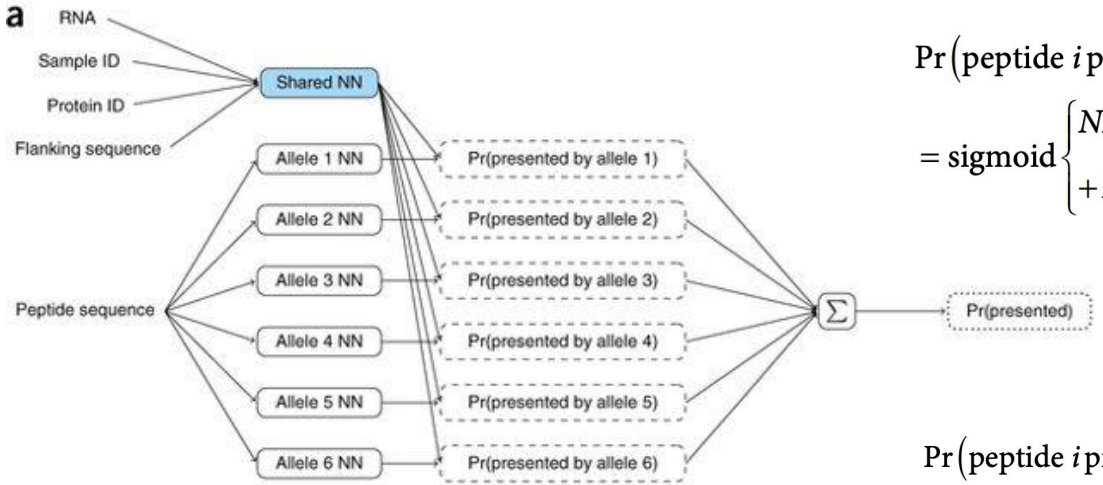


$$\Pr\left(\text{peptide } i \text{ presented by allele } k\right) = \text{sigmoid}\left\{\begin{matrix} NN_k\left(\text{peptide}_i\right) + NN_{\text{flanking}}\left(\text{flanking}_i\right) \\ + NN_{\text{RNA}}\left(\log\left(\text{TPM}_i\right)\right) + \alpha_{\text{sample}(i)} + \beta_{\text{protein(i)}} \end{matrix}\right\}$$

$$\Pr\left(\text{peptide } i \text{ presented}\right) = \sum_{k=1}^{m} a_k^i \times \Pr\left(\text{peptide } i \text{ presented by allele } k\right)$$

https://github.com/nehal96/Deep-Learning-ND-Exercises/blob/master/Deep%20Neural%20Networks/tf-deep-neural-networks.md

Methods - final prediction



$$\Pr\left(\text{peptide } i \text{ presented by allele } k\right) = \text{sigmoid}\left\{\begin{matrix} NN_k\left(\text{peptide}_i\right) + NN_{\text{flanking}}\left(\text{flanking}_i\right) \\ + NN_{\text{RNA}}\left(\log\left(\text{TPM}_i\right)\right) + \alpha_{\text{sample}(i)} + \beta_{\text{protein(i)}} \end{matrix}\right\}$$

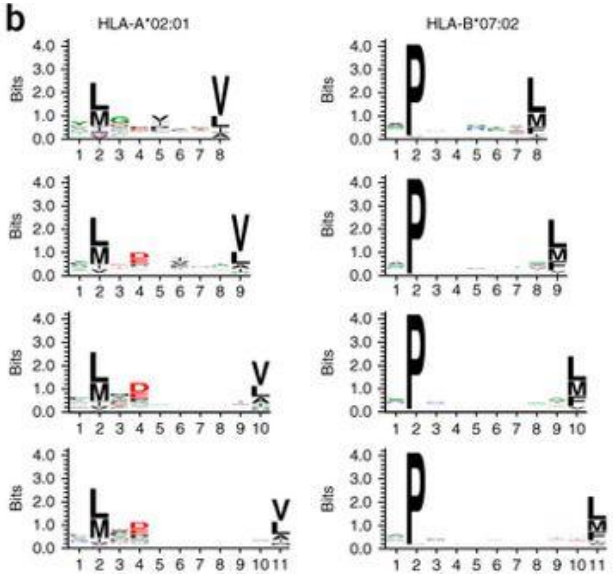
$$\Pr\left(\text{peptide } i \text{ presented}\right) = \sum_{k=1}^{m} a_k^i \times \Pr\left(\text{peptide } i \text{ presented by allele } k\right)$$

https://github.com/nehal96/Deep-Learning-ND-Exercises/blob/master/Deep%20Neural%20Networks/tf-deep-neural-networks.md

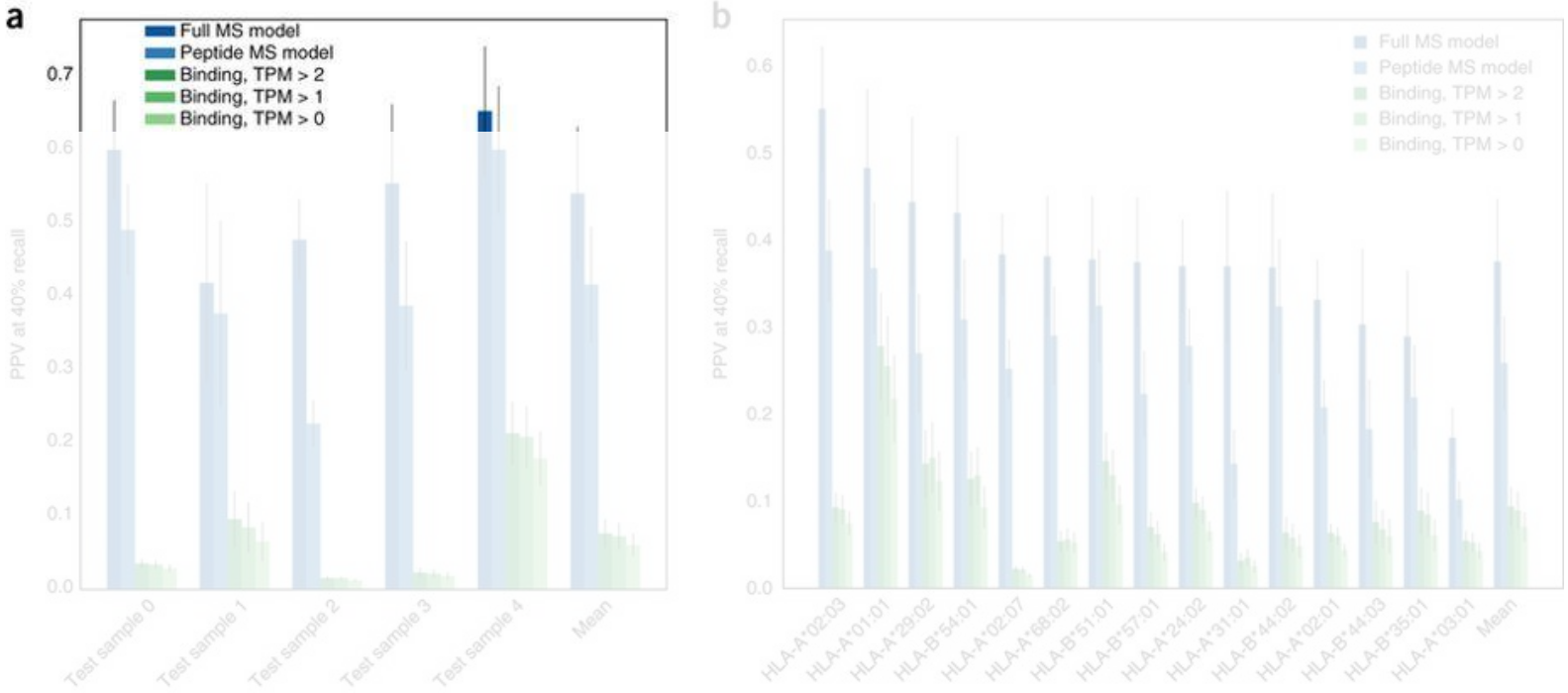
Methods - model training

$$\text{Loss}(i) = -\log\left(\text{Bernoulli}\left(y_i \mid \Pr\left(\text{peptide } i \text{ presented}\right)\right)\right)$$

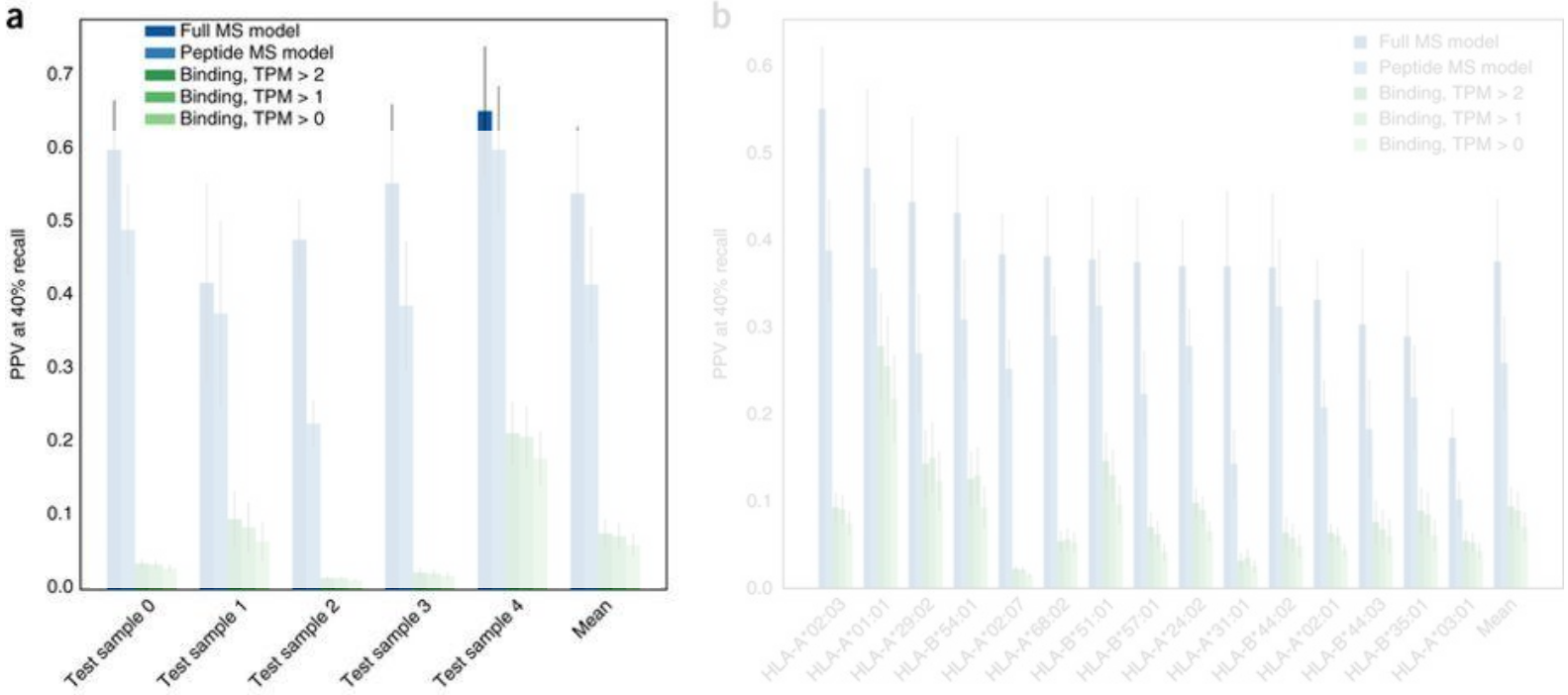
Results - different HLA indeed learns some pattern in the peptide sequence



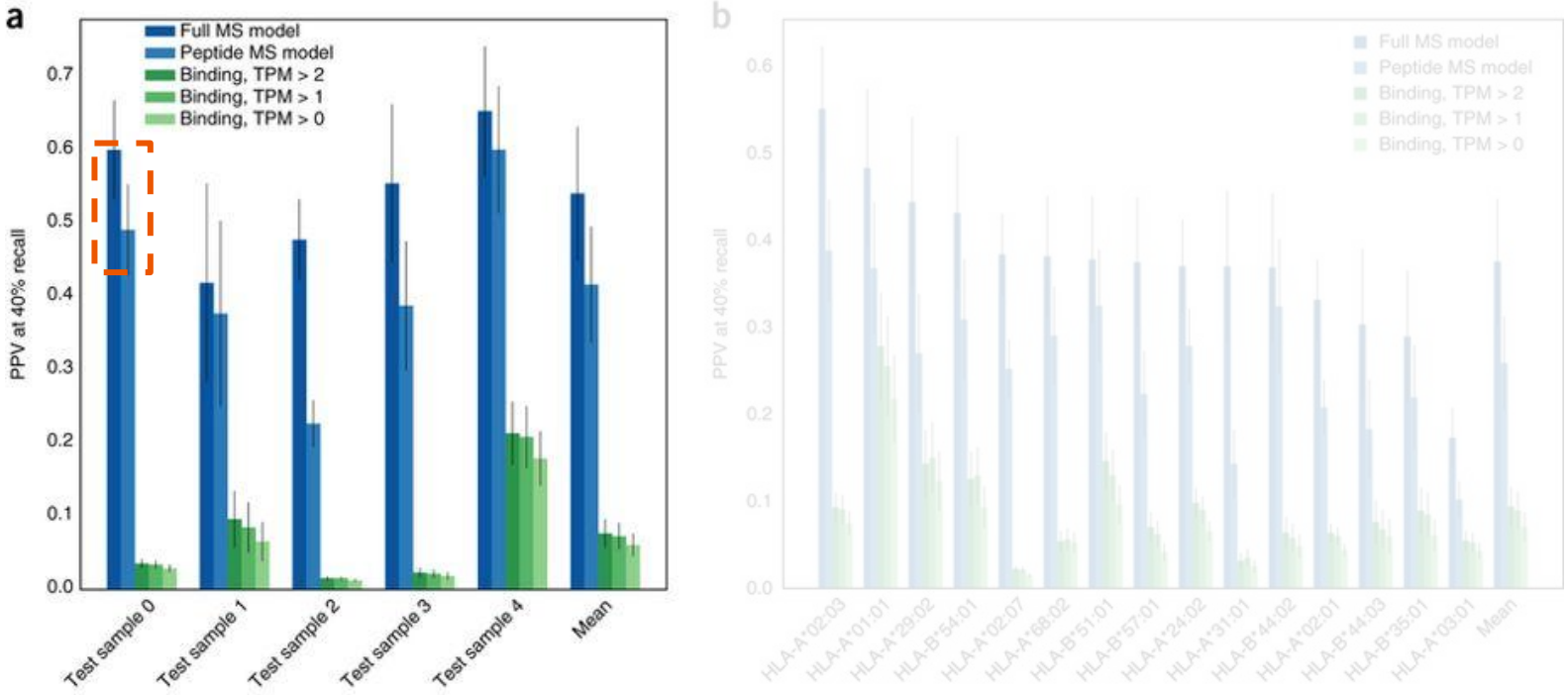
Results - benchmarks



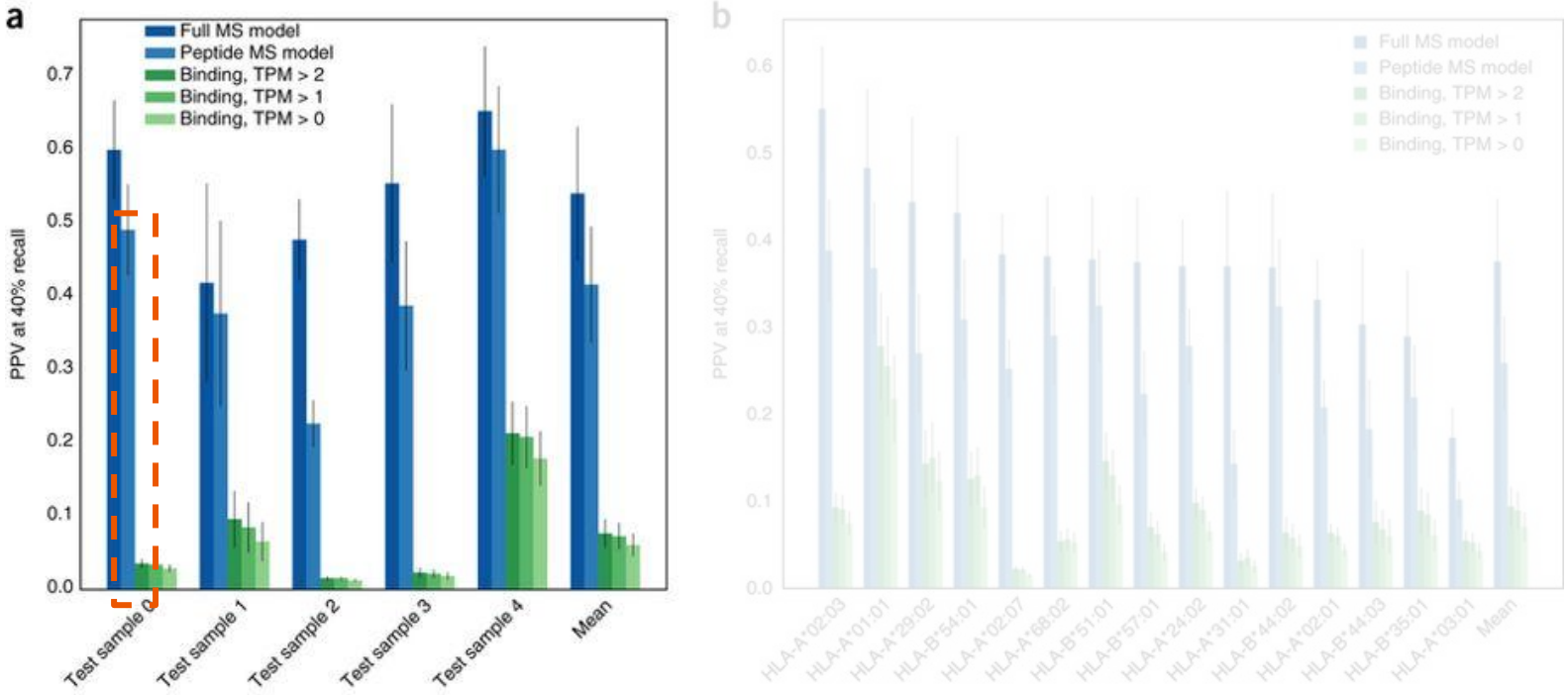
Results - evaluation metric



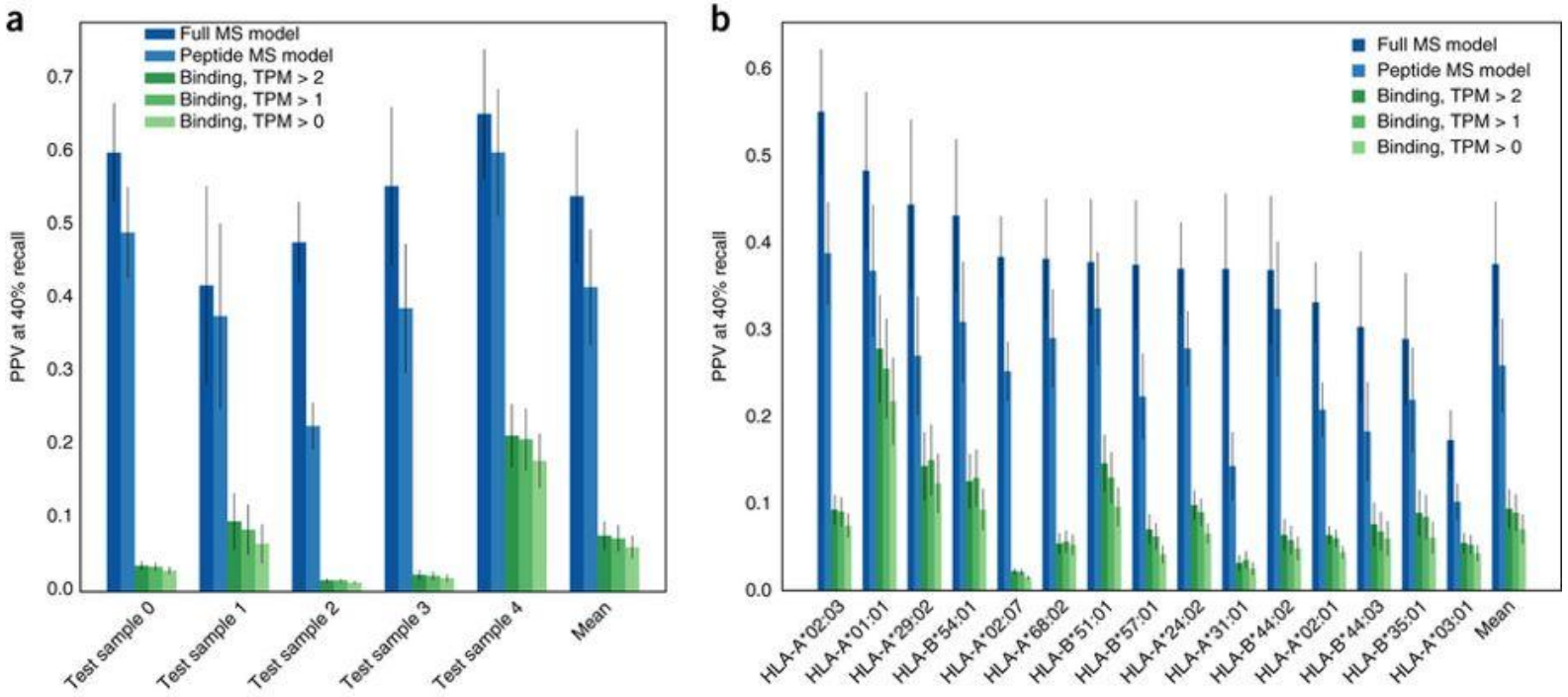
Results - peptide presentation prediction performance



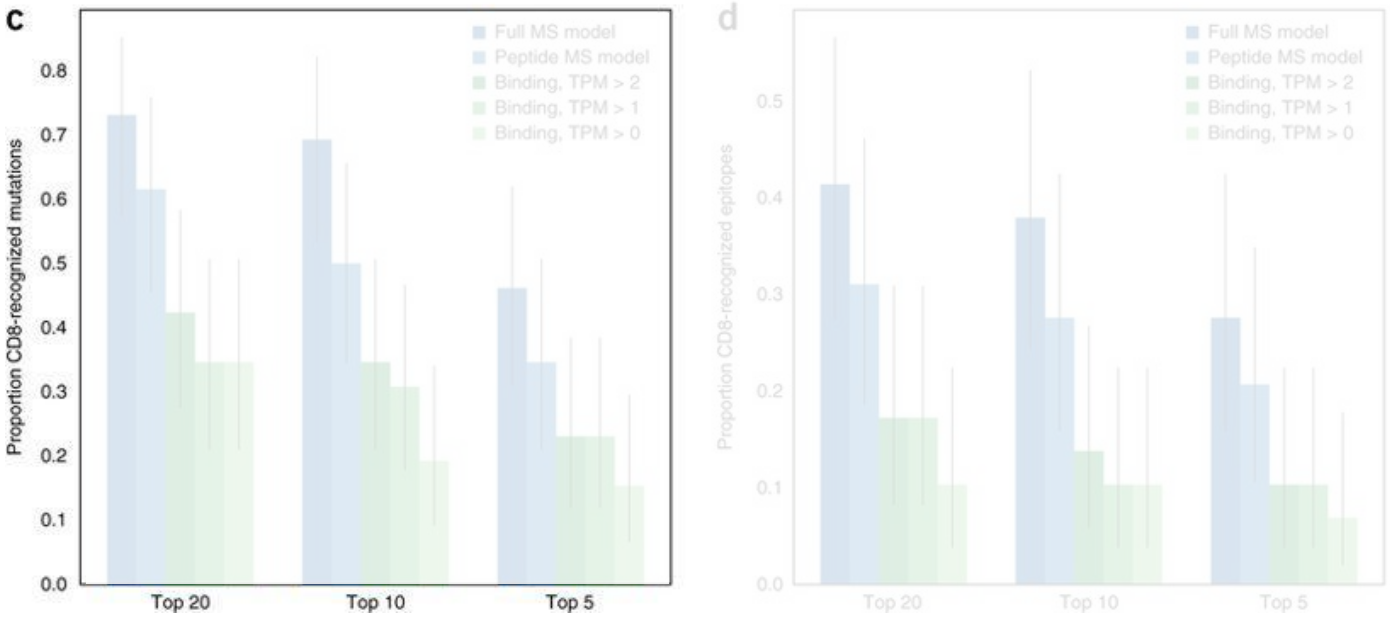
Results - it performs very well



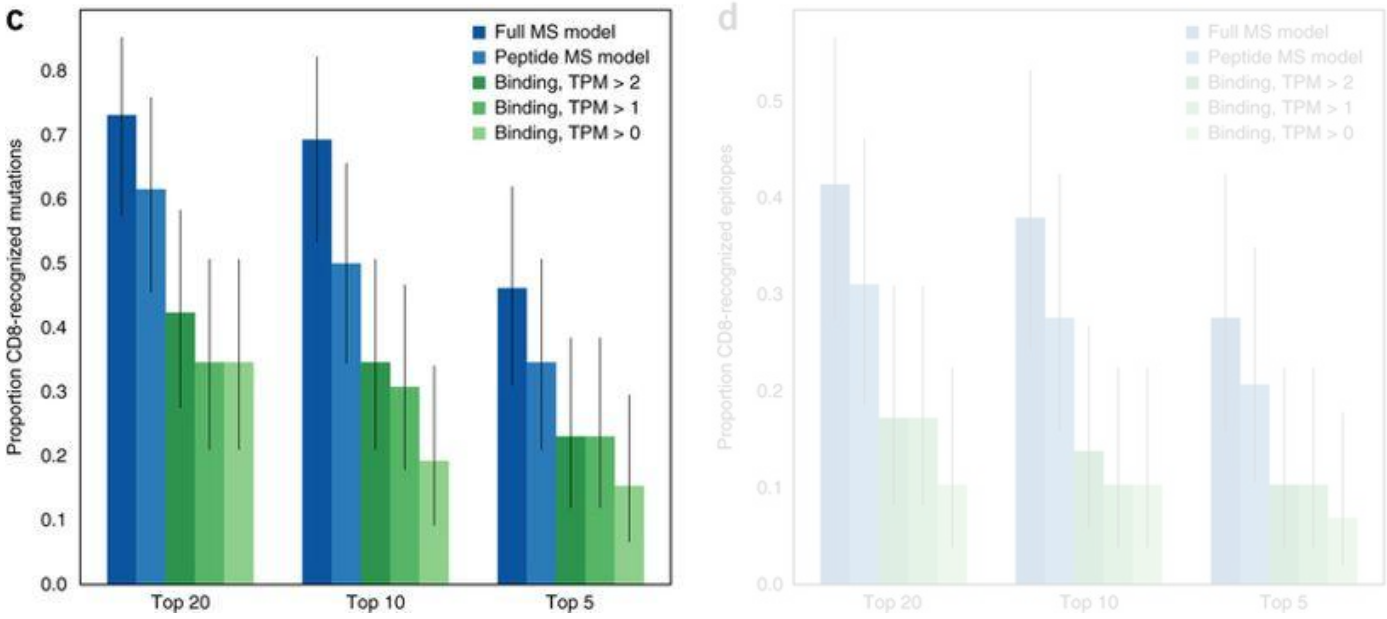
Results - it performs very well



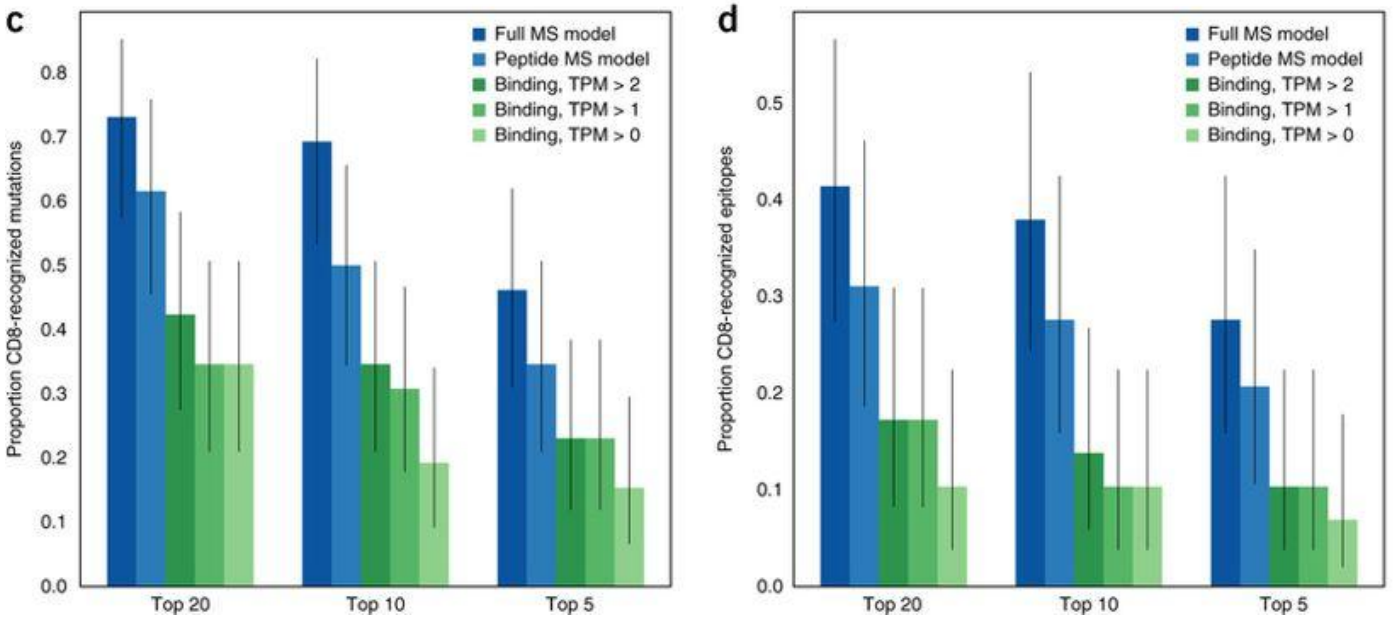
Results - actual t-cell recognition



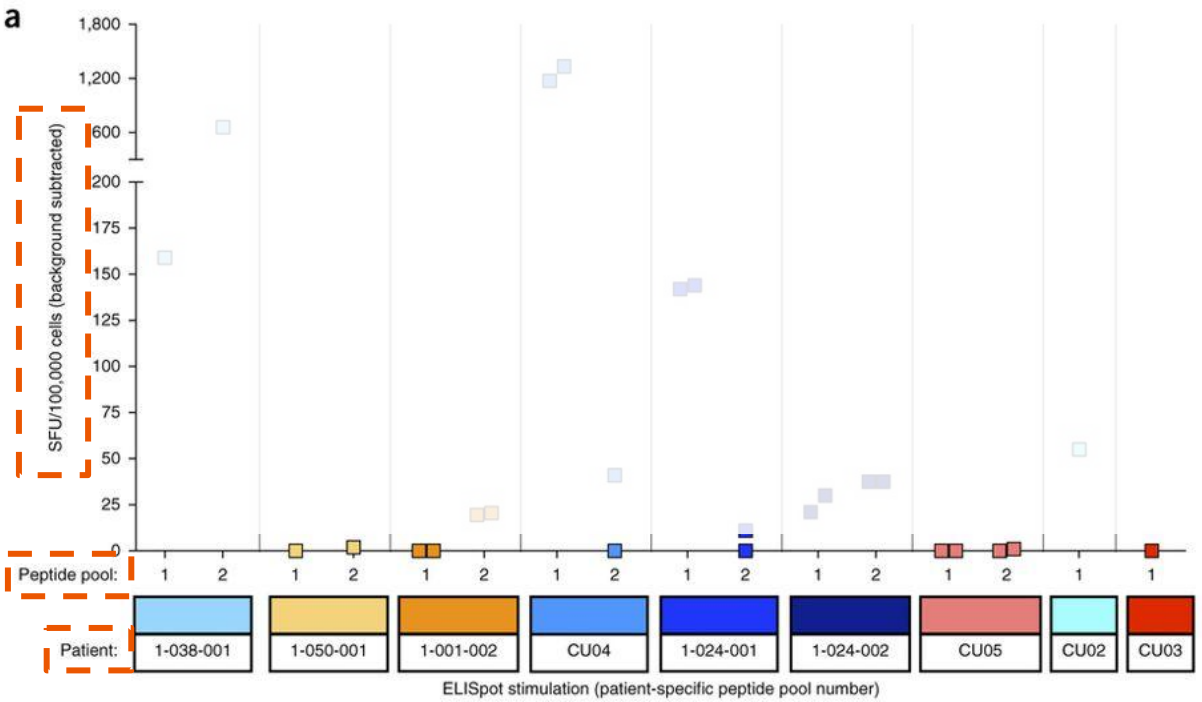
Results - it against performs very well



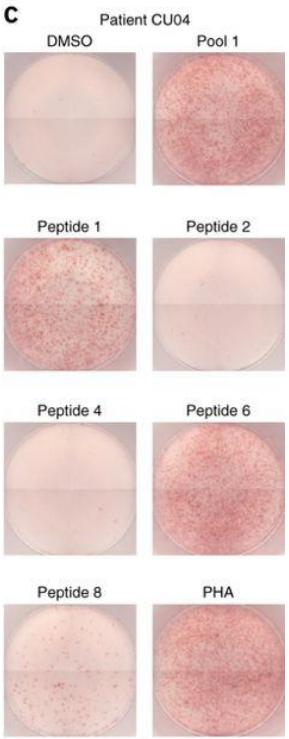
Results - it against performs very well



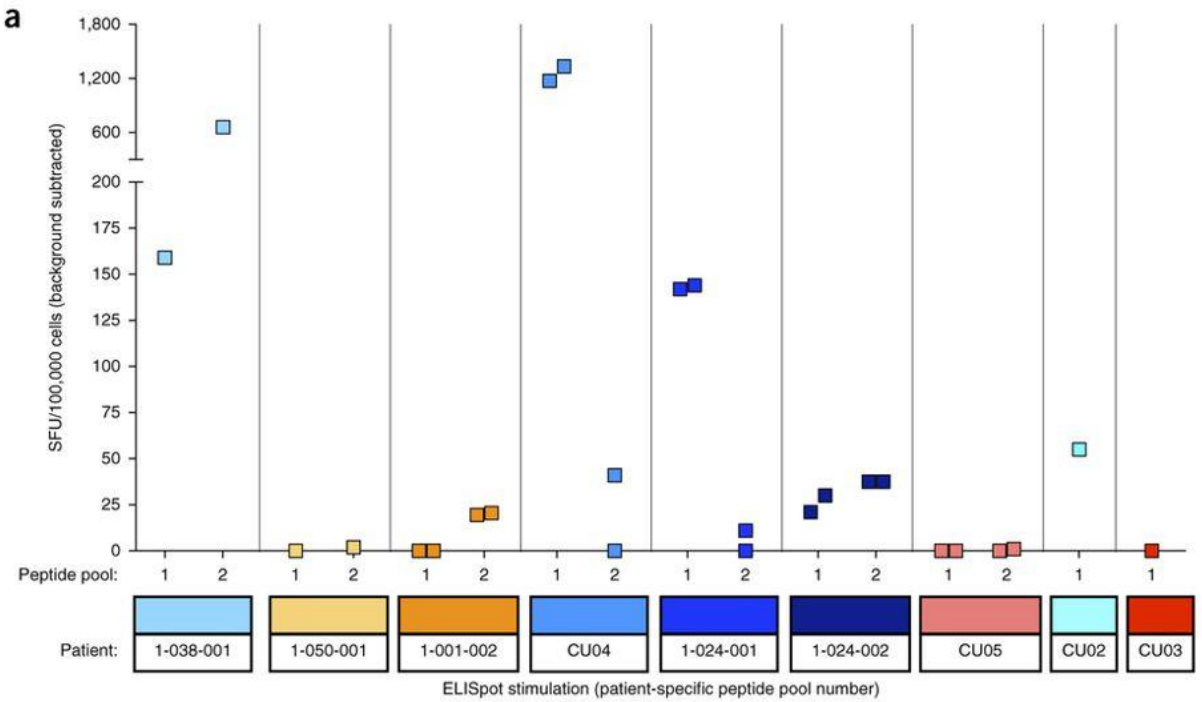
Results - neoantigen identification in real cancer patient



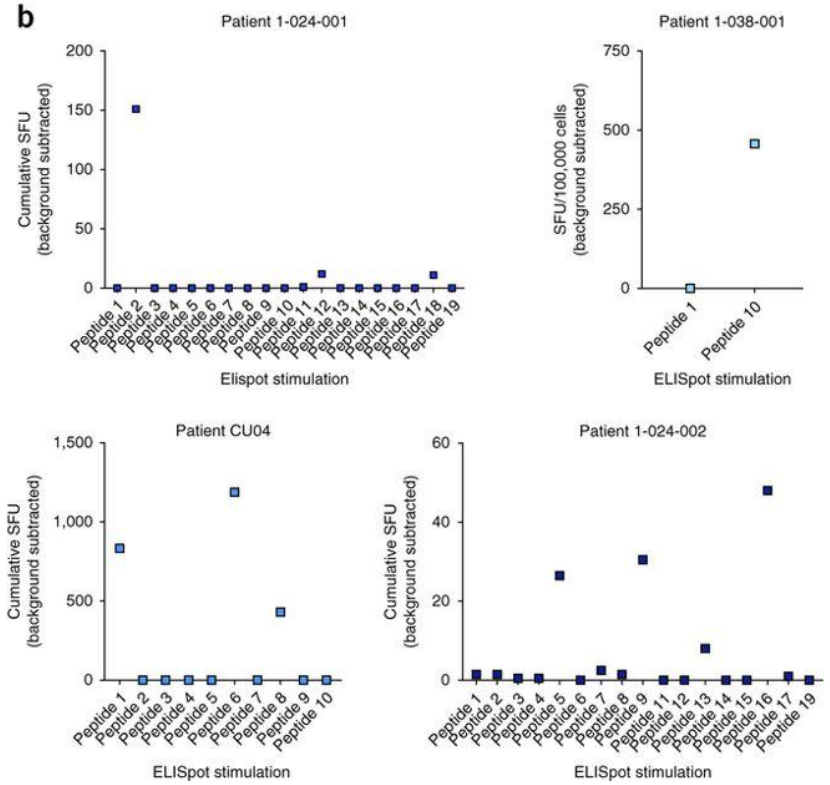
Results - surface formation units



Results - there are some promising signals



Results - there are some promising signals



Conclusion

- Incorporating peptide-extrinsic features helps
- Better model the peptide sequence with different alleles type helps

- Lot's of work go into the data collection (clean up, clinical experiments etc.) to make this an interesting paper.

- Extend this work to HLA Class II binding
- Consider the TCR binding or T-Cell precursor information (i.e. modeling more information in the pathway)
- Better neural network model?

Conclusion

- Incorporating peptide-extrinsic features helps
- Better model the peptide sequence with different alleles type helps

- Lot's of work go into the data collection (clean up, clinical experiments etc.) to make this an interesting paper.

- Extend this work to HLA Class II binding
- Consider the TCR binding or T-Cell precursor information (i.e. modeling more information in the pathway)
- Better neural network model?

Conclusion

- Incorporating peptide-extrinsic features helps
- Better model the peptide sequence with different alleles type helps

- Lot's of work go into the data collection (clean up, clinical experiments etc.) to make this an interesting paper.

- Extend this work to HLA Class II binding
- Consider the TCR binding or T-Cell precursor information (i.e. modeling more information in the pathway)
- Better neural network model?

Q&A