Computational prediction of MHC class I tumorspecific antigens

Anton Alexandrov

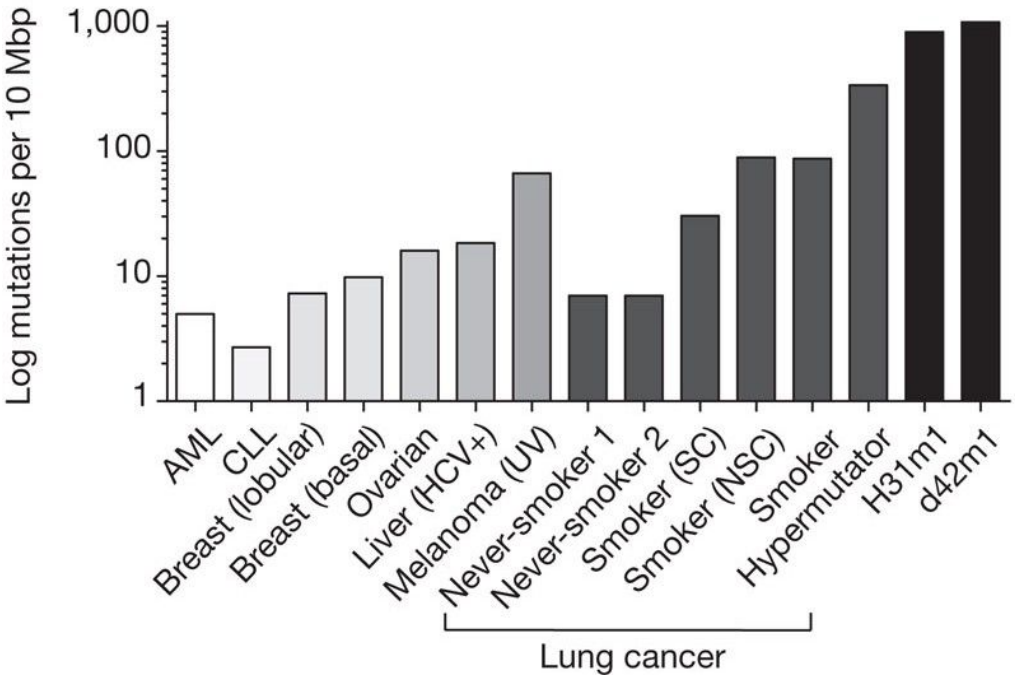
ITMO University, Saint Petersburg, Russia, `alantbox@gmail.com`

Maxim Artyomov

Washinton University in St Louis, `martyomov@pathology.wustl.edu`

One of the most promising approaches for cancer treatment is immunotherapy. It is possible to direct the immune system to attack cancer cells. This can be done because of the presence of foreign proteins inside cancer cells. The immune system recognizes those foreign proteins and kills the cells containing them.

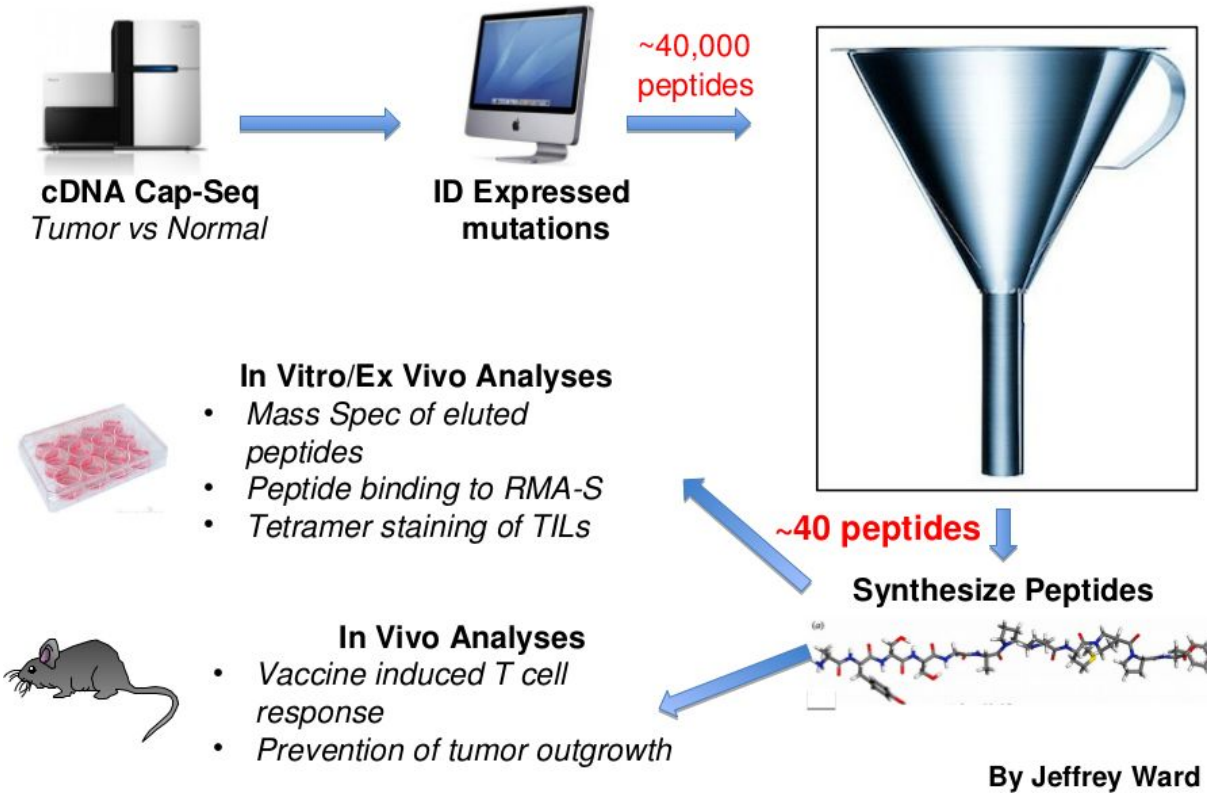
Why are there foreign proteins inside cancer cells? This is because cancer genome is slightly different from the genome of the host organism. The more mutations cancer has the more foreign proteins its cells express. The mutational landscape of different cancer is a well studied topic. Statistics for selected types of cancers is shown on picture 1.



Picture 1. Mutation density in cancer cells

The chance of finding an immunogenic peptide among all neutral foreign peptides for a mutation-rich cancer is higher than that for a cancer with fewer mutations. This is why we based our study on two murine cancer cell lines: H31m1 and d42m1.

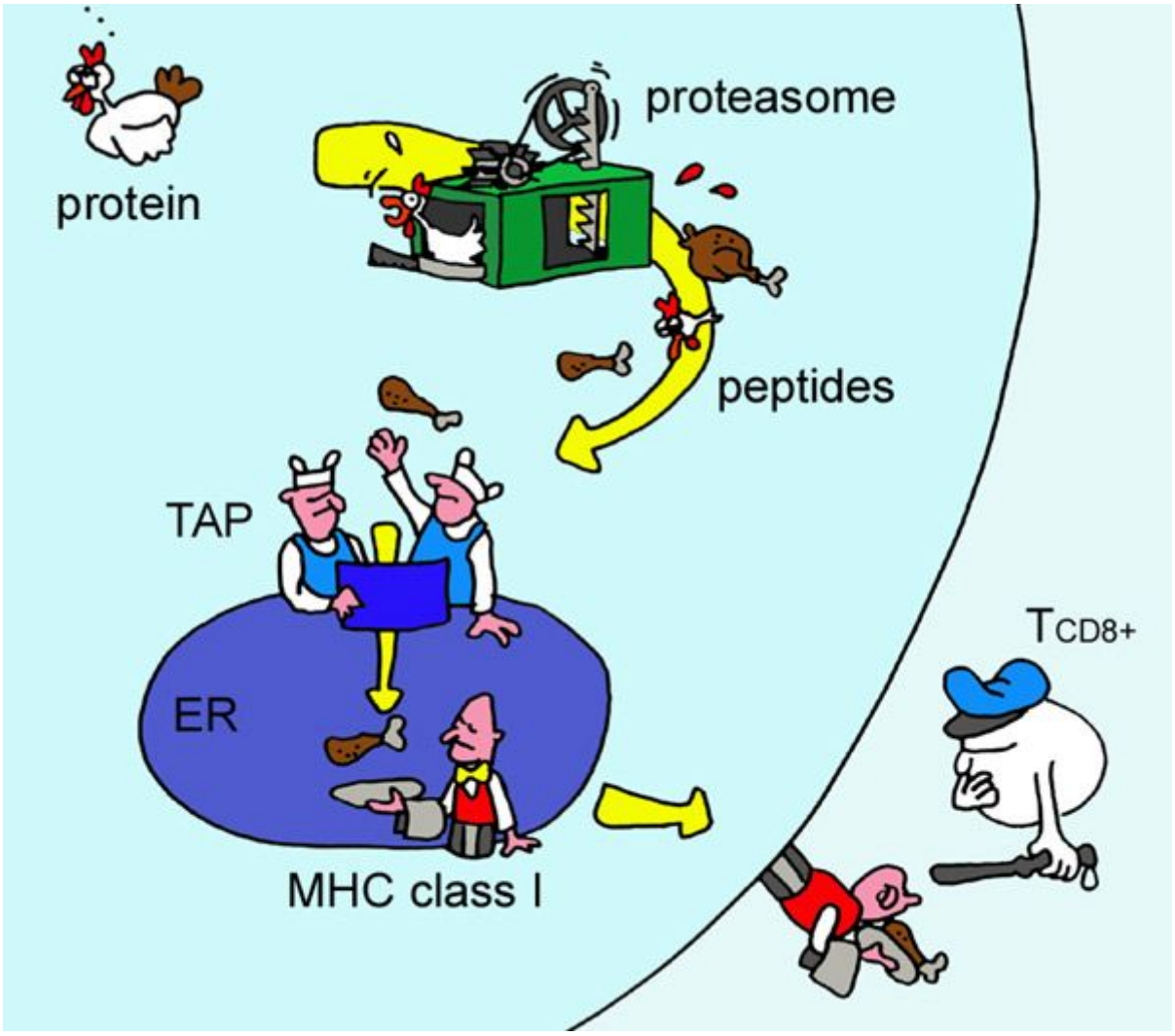
The selected tumors have about 1000 coding mutations each, which leads to tens of thousands of different peptides, and each of them is potentially immunogenic. Assessing each peptide with in-vivo or in-vitro methods is expensive and takes a lot of time. Therefore there is a need for a computational method capable of ruling out definitely non-immunogenic peptides (picture 2).



Picture 2. Immunogenic epitope prediction pipeline

After the number of potentially immunogenic peptides has been reduced to a small group of peptides all of those can be studied in-vitro or in-vivo and the best vaccine can be made.

There are several factors that contribute to the final peptide presentation landscape of the cell (picture 3).



Picture 3. MHC class I epitope presentation pathway

The first step a peptide goes through is the proteasome. Here the peptide is cut into short peptides (8-11 amino acids long). The second step is the MHC class I molecule which binds the peptide (or not) and presents it on the cell surface. The final step is the CD8+ cell that kills the cells showing foreign peptides on their surface.

Computational prediction of the proteasome work is a simple process and there is one well-known algorithm for that (NetChop [1]) which can be used to calculate the processing score for a given peptide. However, prediction of binding affinity for a given peptide can present some difficulties. There are many methods (NetMHCpan [2], ANN[3], SMM[4]), and each of

them takes into account a particular trait of the peptide necessary for binding the MHC molecule (anchor residue, peptide length, etc). Therefore none of the methods performs well in all cases. In order to combine the advantages of each method without losing much because of the disadvantages we use all of the specified methods and calculate median affinity value for a given peptide.

We also introduce a new value – **neoepitope ratio**, which is equal to the mutant peptide median affinity divided by the median affinity of its wild-type counterpart. This value shows how much more immunogenic the mutant peptide is than its wild-type counterpart. To summarize, the algorithm is as follows:

1. For a given list of cancer mutations leave out non-coding mutations.
2. Construct a list of wild-type–mutant peptides of length 8-11 aa.
3. Calculate the processing score for each peptide.
4. Calculate the median affinity value for each peptide.
5. Calculate the neoepitope ratio for each peptide pair.

For each of the calculated values we determined the threshold value that allows us to achieve maximal sensitivity without high increase in false-positives fraction. This approach has been used in several projects at Washington University in St Louis, one of them has already led to a publication [5].

References:

1. M. Nielsen, C. Lundegaard, O. Lund, and C. Kesmir. The role of the proteasome in generating cytotoxic T cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. Immunogenetics., 57(1-2):33-41, 2005.

2. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 61:1-13.

3. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, and Nielsen M. 2008. NetMHC-3.0: Accurate web accessible predictions of Human, Mouse, and Monkey MHC class I affinities for peptides of length 8-11. NAR 36:W509-512.

4. Peters B, Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6:132.

5. Gubin et al. 2014. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature, 515, pp. 577–581 (27 November 2014)