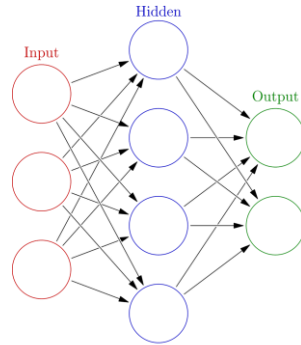




# Designing cancer vaccines with deep learning



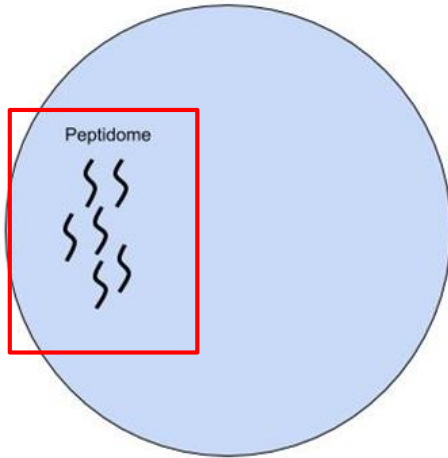
Mikhail Ignatov,  
Applied  
Mathematics  
and Statistics  
PhD candidate

# Why doesn't the immune system treat cancer?

- 1) Tumor microenvironment suppresses activity of immune cells
- 2) Tumor cells are very similar to healthy cells

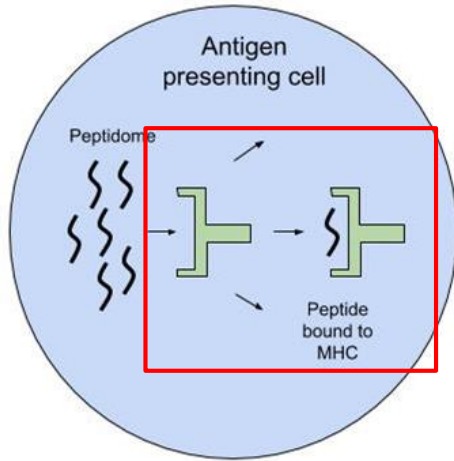
We can force the immune system to target tumor cells using  
cancer immunotherapy

# How to target tumors?



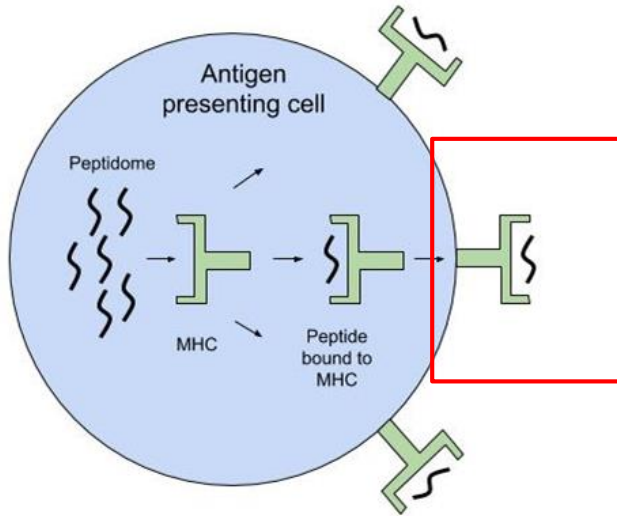
- 1) Proteins in a cell are sliced into small pieces - peptides

# How to target tumors?



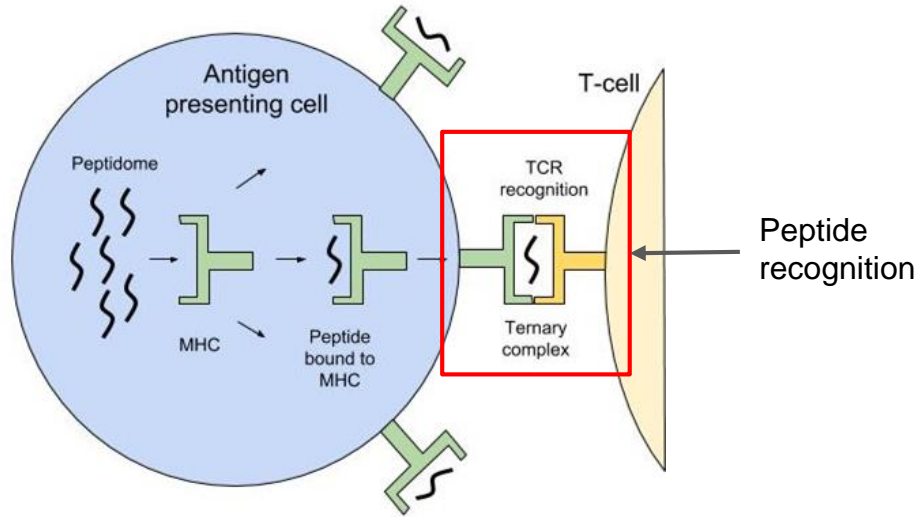
- 1) Proteins in a cell are sliced into small pieces - peptides
- 1) Peptides are loaded onto MHC molecules

# How to target tumors?



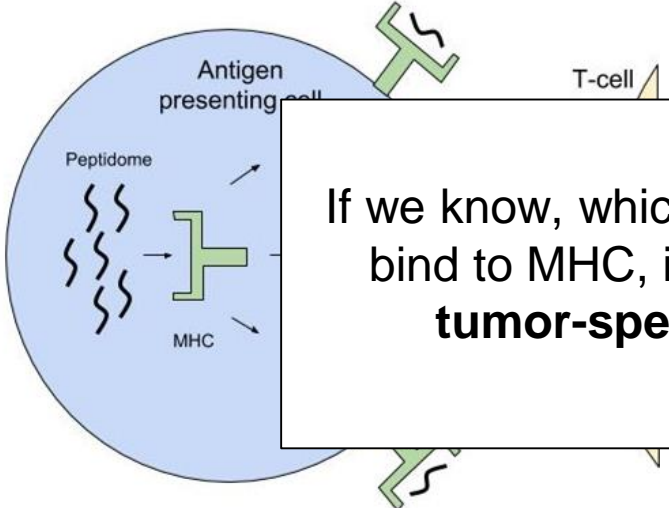
- 1) Proteins in a cell are sliced into small pieces - peptides
- 1) Peptides are loaded onto MHC molecules
- 1) MHC transport peptides to cell membrane and expose them to T-cells

# How to target tumors?



- 1) Proteins in a cell are sliced into small pieces - peptides
- 1) Peptides are loaded onto MHC molecules
- 1) MHC transport peptides to cell membrane and expose them to T-cells
- 1) Upon recognition of the peptides T-cell can destroy the antigen presenting cell

# How to target tumors?



If we know, which tumor derived peptides bind to MHC, it is possible to develop **tumor-specific T-cells. But ...**

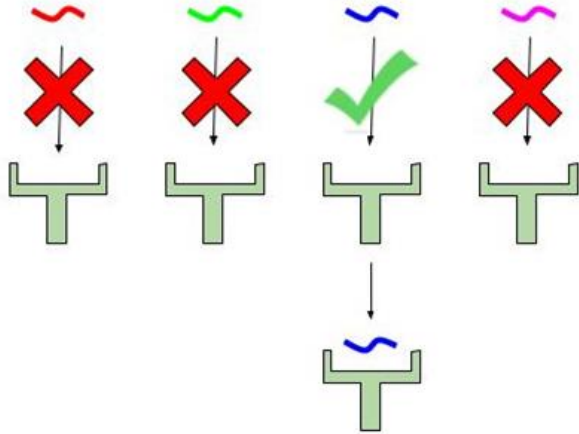
1) Proteins in a cell are sliced into small pieces - peptides

loaded onto MHC

port peptides to cell and expose them to T-cells

1) Upon recognition of the peptides T-cell can destroy the antigen presenting cell

# The problem

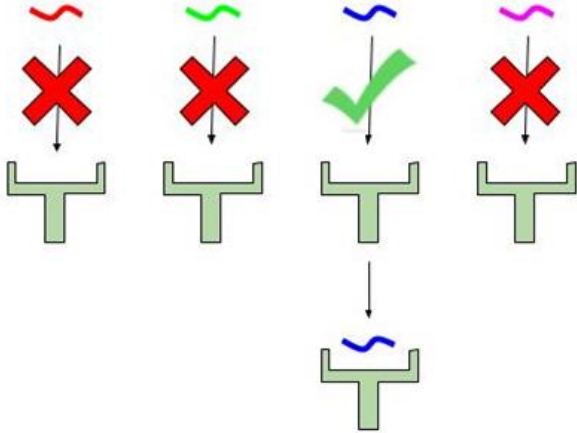


There are millions of different peptides in the cell.

**How to predict, which of them bind to MHC and become presented to the immune system?**



# The problem

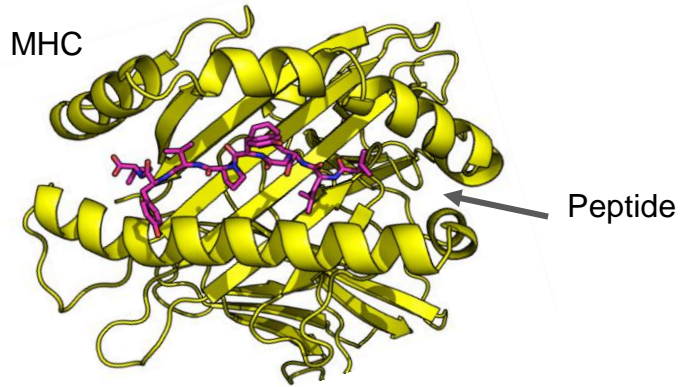


There are millions of different peptides in the cell.

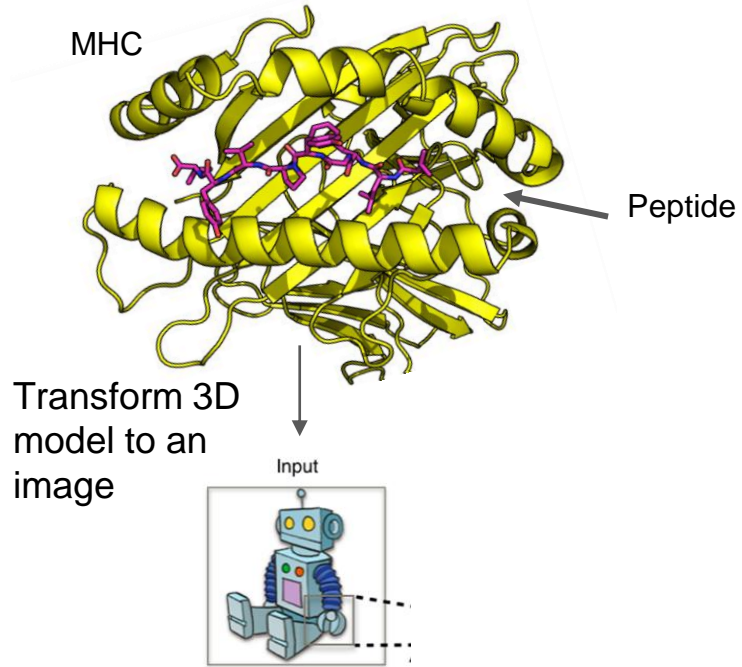
**How to predict, which of them bind to MHC and become presented to the immune system?**

Using Machine Learning

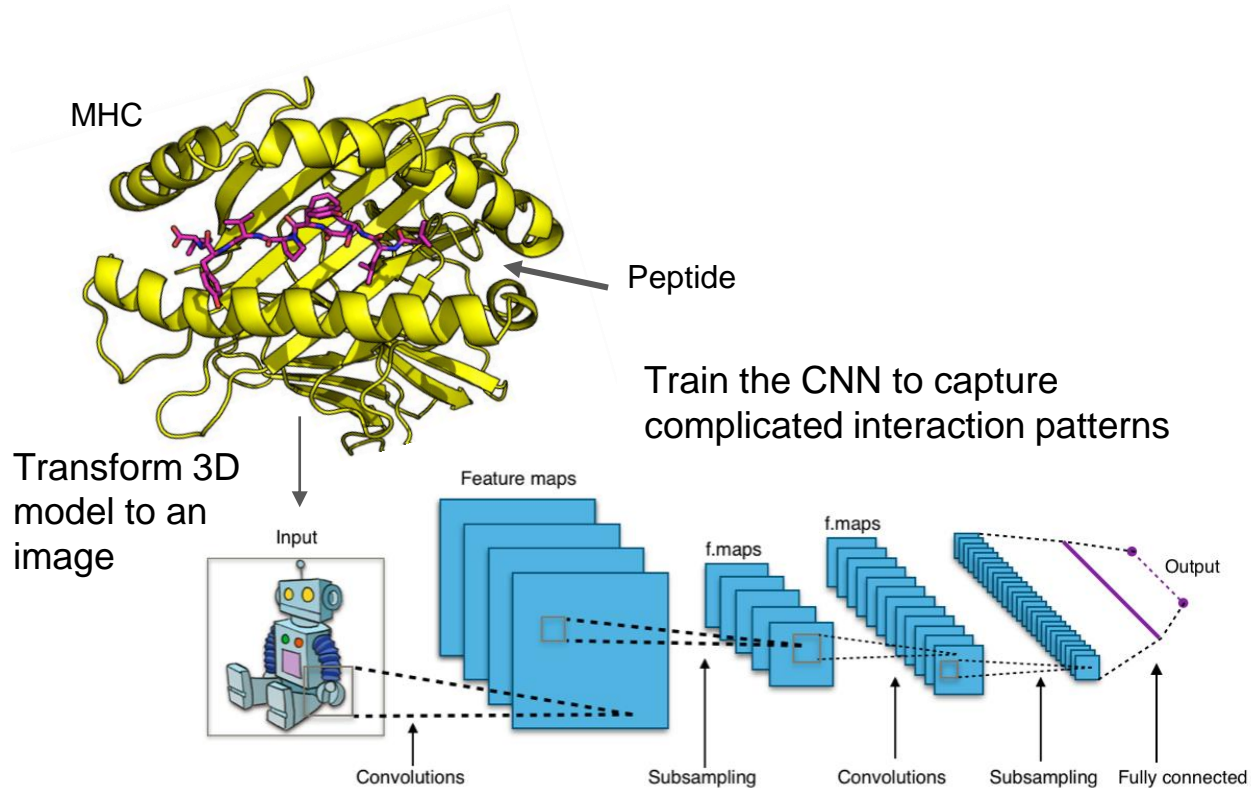
# Convolutional Neural Nets for binding prediction



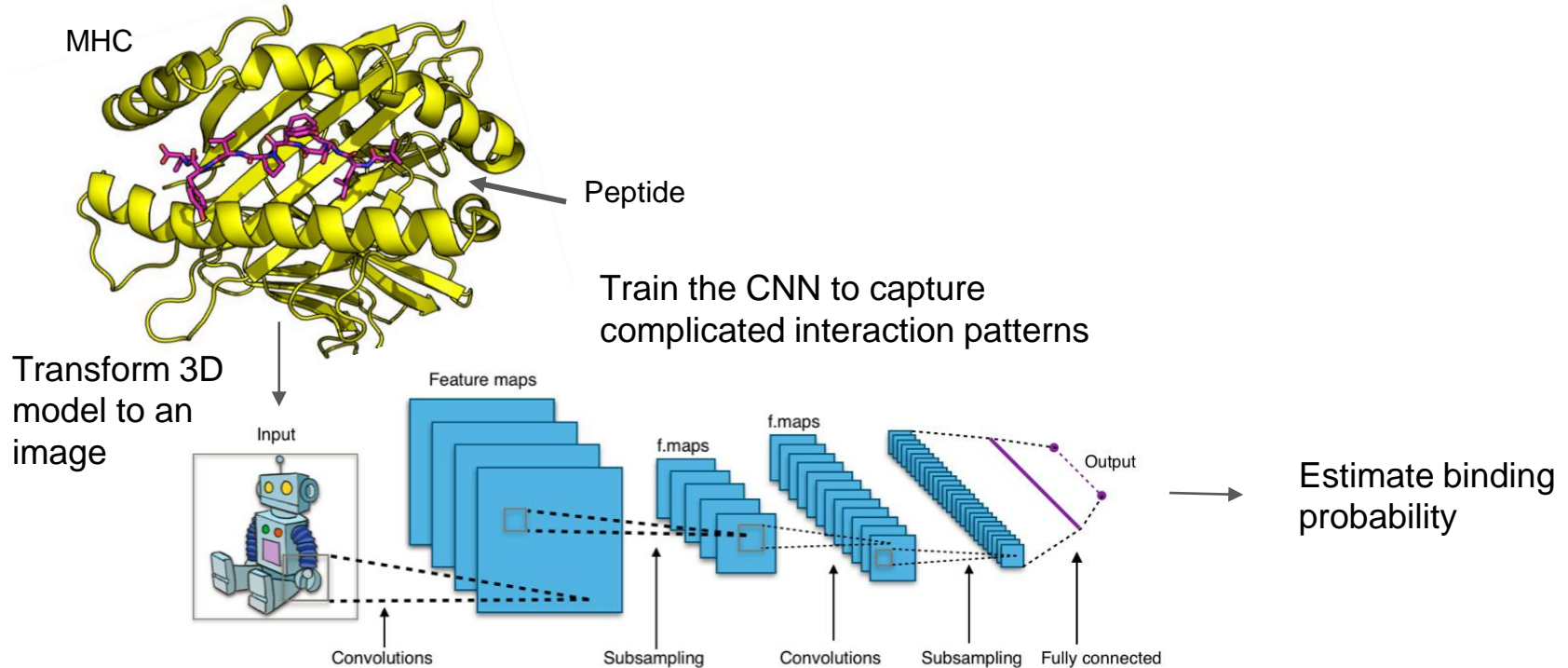
# Convolutional Neural Nets for binding prediction



# Convolutional Neural Nets for binding prediction



# Convolutional Neural Nets for binding prediction



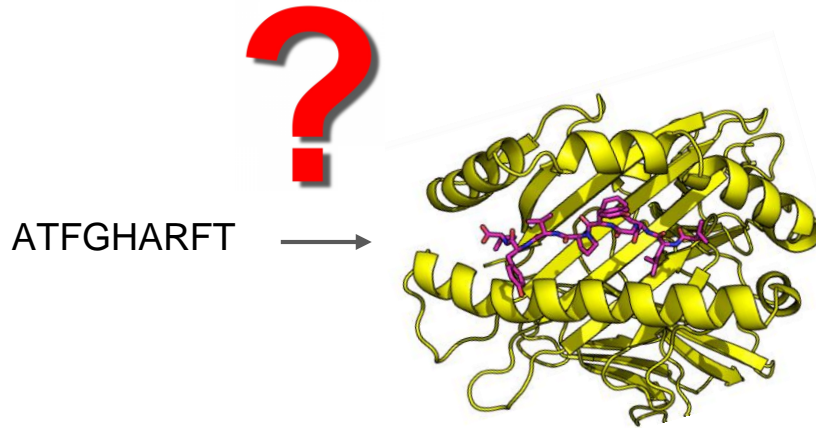
# Convolutional Neural Nets for binding prediction

## Why CNNs?

- 1) CNNs are highly relevant for image recognition tasks
- 1) A large number of frameworks exists for implementing CNNs, because of their high popularity
- 1) All the peptide-MHC interactions are highly similar in orientation, terminis are fixed
- 1) Sufficient data for training: ~400000 (non)interacting pairs
- 1) All current prediction tools rely solely on sequence data, without taking structural information into account. This results in very poor predictions for less frequent MHC variants and as a consequence in ineffective treatment

# How to create good quality models?

Most of the training data exists only in sequence form  
We need to be able to properly model peptide-MHC interaction



# Model ranking using SVM

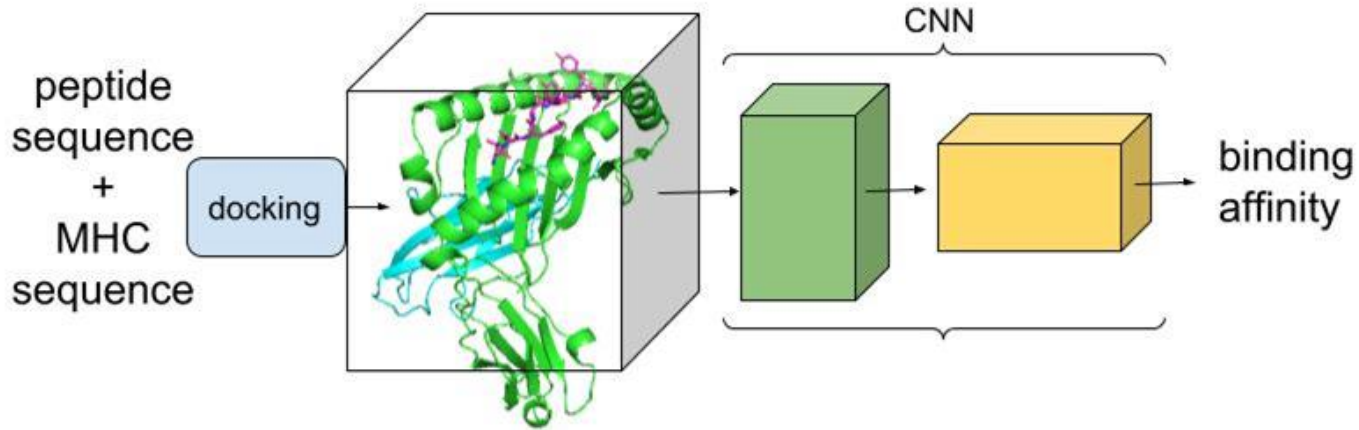
Modeling protocol:

- 1) Randomly sample multiple peptide conformations in MHC binding site
- 2) Minimize molecular mechanics energy function of each model
- 3) Sort models by similarity to the real conformation (RMSD)
- 4) Compute a large number of chemical features for each model
- 5) Train energy function specific for peptide-MHC interaction using SVM



# Pipeline

- 1) Model MHC-peptide interaction using SVM-based energy function
- 2) Predict binding score of the created model



# Timeline

1st year:

## Docking protocol design steps

1. Set up the testing framework for development of energy function (currently in development)
2. Train a number of SVM models using different feature sets (Rosetta energy terms)
3. Select the best SVM model for peptide-MHC docking

## CNN design steps

1. Prepare the data, create MHC-peptide models, choose training and testing sets
2. Train and test multiple CNNs with different architectures, pick the best one
3. Test the resulting CNN model on external datasets

2nd year:

Study the possibility of adding patient's T-cell repertoire.  
Learn to model T-cell interactions with peptide-MHC complex  
and introduce it into the prediction pipeline

Thank you for attention