

Artificial Intelligence in Medicine:
next steps,
October 17th 2024

Unraveling the network signatures of oncogenicity in virus-human protein-protein interactions

Francesco Zambelli,
Vera Pancaldi,
Manlio De Domenico



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

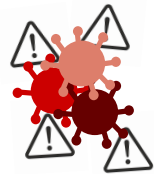


CoMuNe lab
COMPLEX MULTILAYER NETWORKS

The Problem



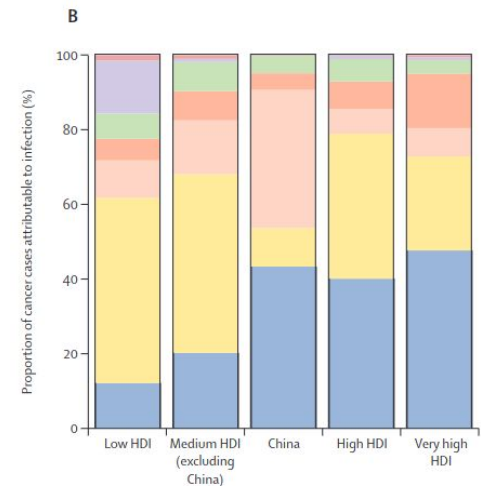
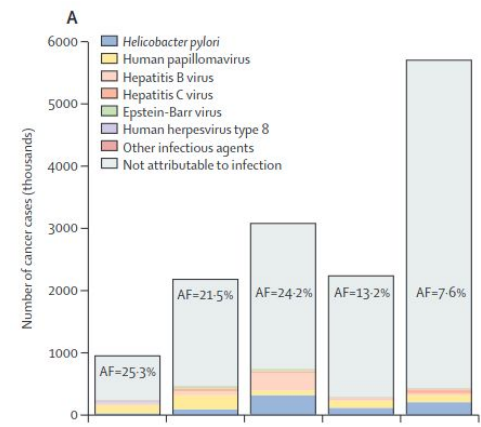
Infectious diseases are increasingly frequent



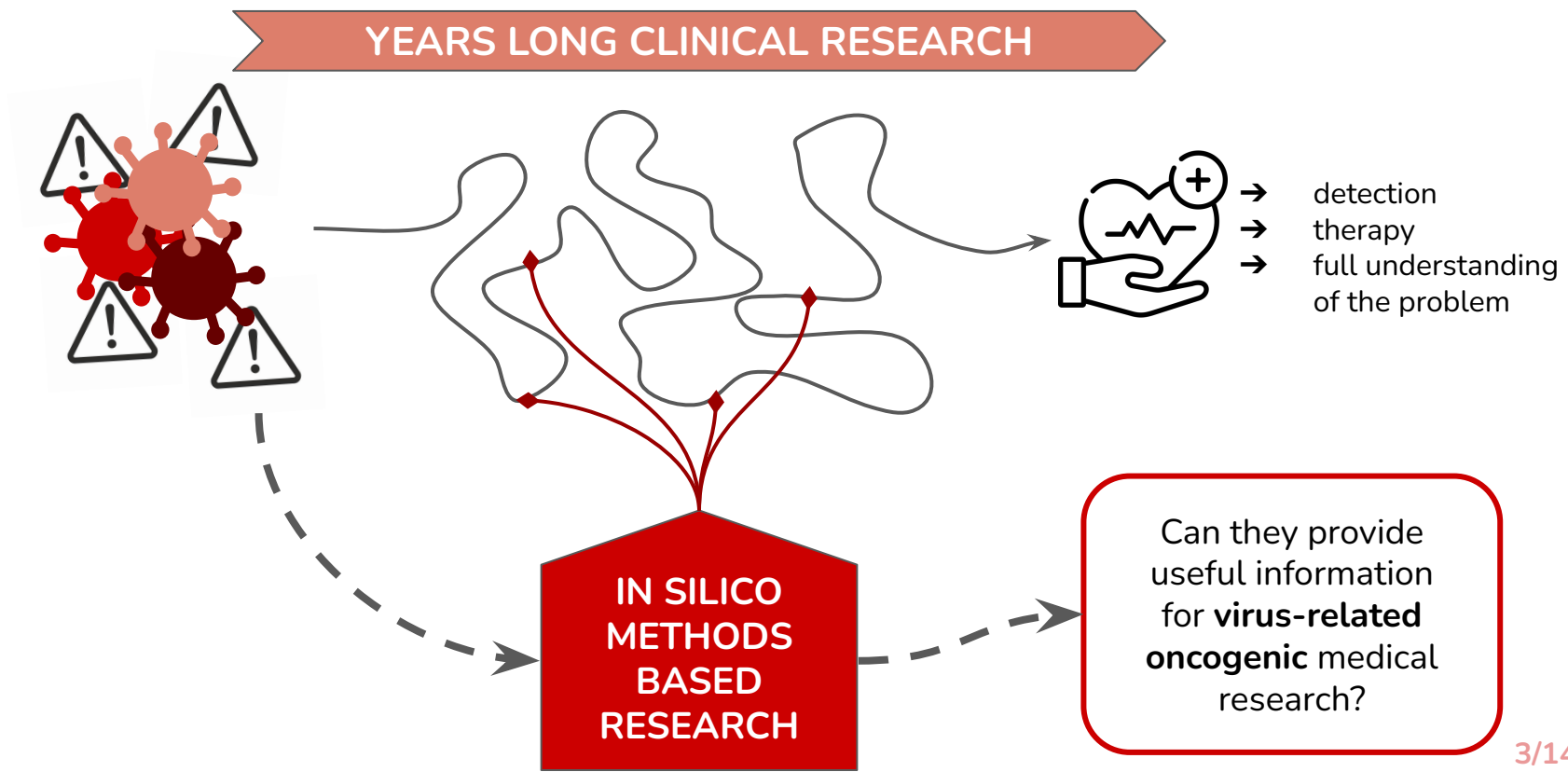
Virus-associated long term effects



Focus on increased cancer risk
ONCOGENICITY



The Question



Methods

Network Medicine



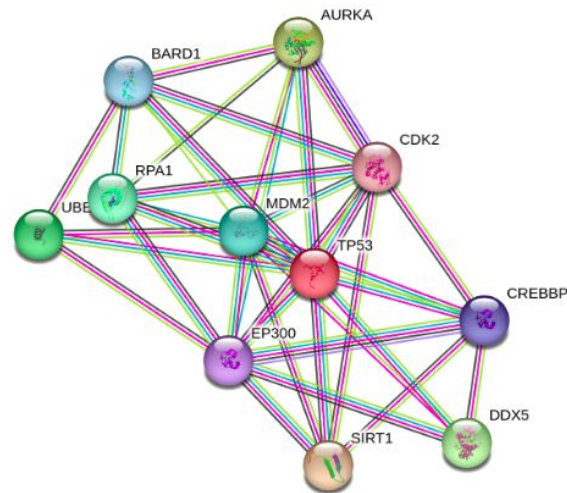
Physical model for the virus-human interaction



Protein-protein interaction networks



1. Classification
2. Relevant features detection



Edges:

Edges represent protein-protein associations
associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.

Known Interactions

- from curated databases
- experimentally determined

Predicted Interactions

- gene neighborhood
- gene fusions
- gene co-occurrence

Others

- textmining
- co-expression
- protein homology

Methods

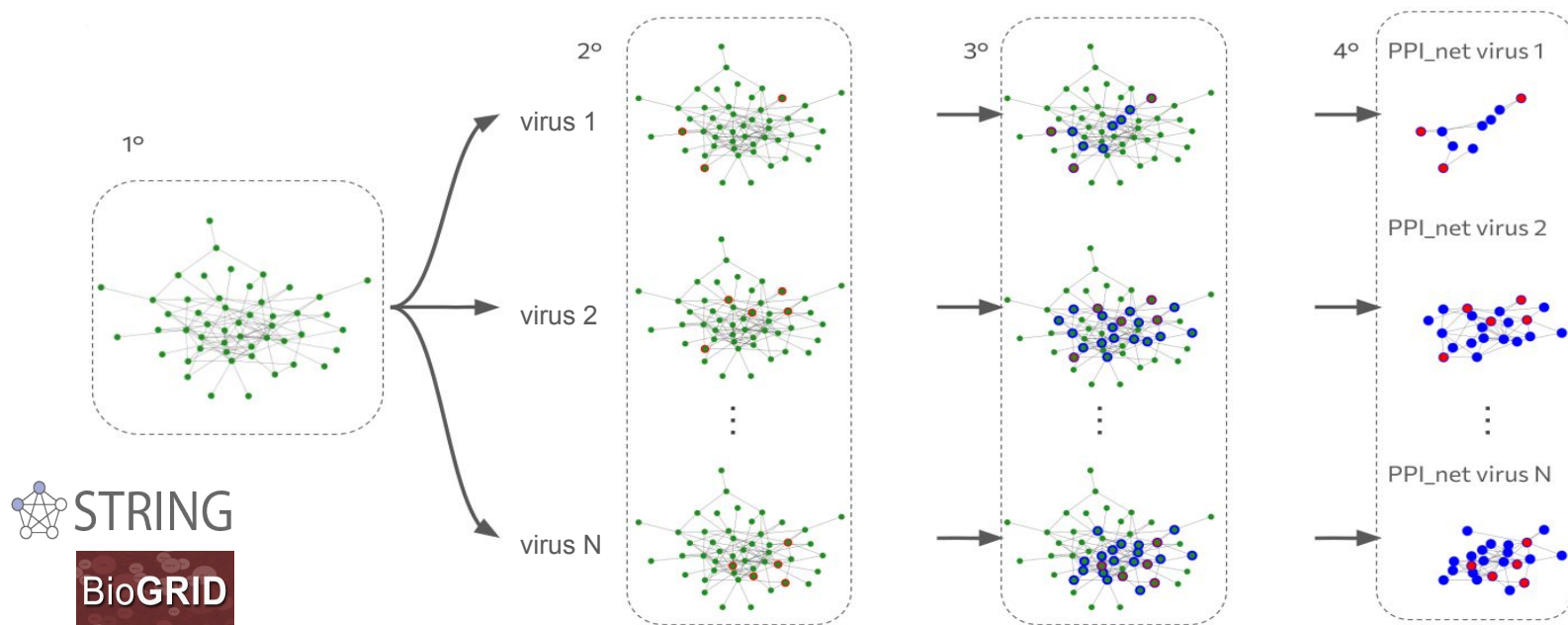
Dataset

Complete human PPI network

Directly targeted proteins

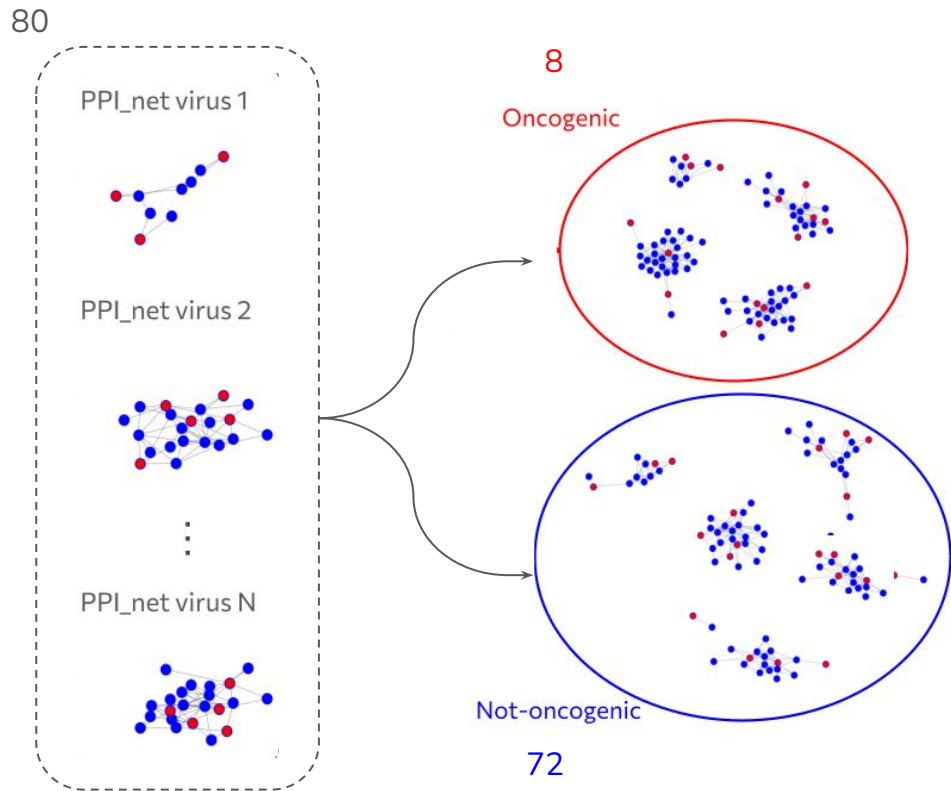
1st order neighbors

80 virus-host interaction PPI networks



Methods

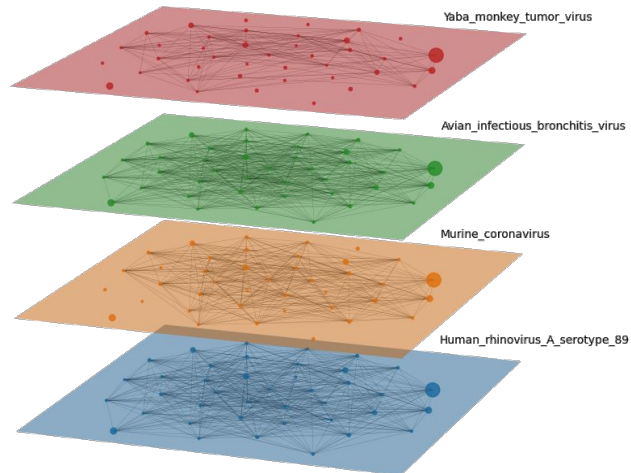
Dataset



- Highlight the common features
- Look for differences to use for classification

Methods

Multilayer approach



INTERPRETATION:

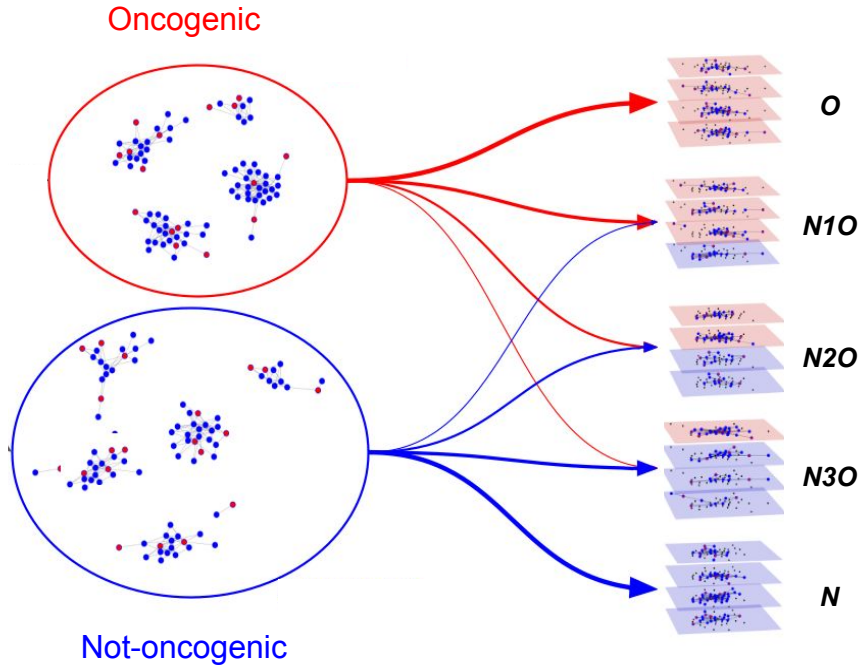
different way in which the human proteome can be influence by viral infections

ADVANTAGES:

focus on shared properties

Methods

Multilayer approach, combination sets



COMBINATION SETS

- Combine oncogenic and not-oncogenic layers in different proportion
- Focus on the “oncogenicity” feature
- Creating **many samples** to perform statistics
- Look for:
 - **clear distinction** of the extreme cases
 - **progressive transition** between them

Results

Topological features

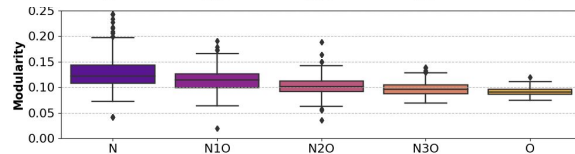
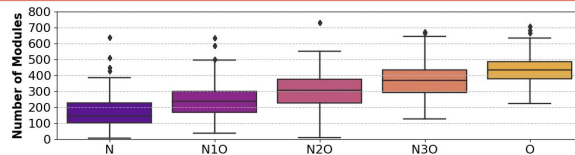
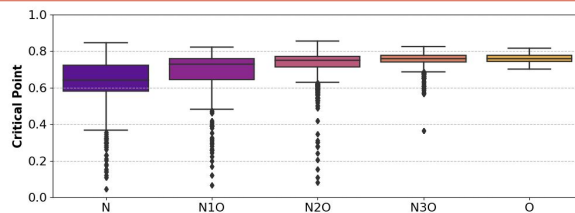
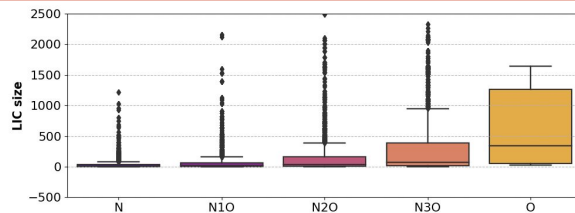
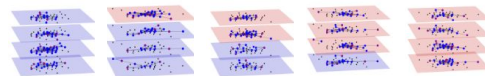
TOPOLOGICAL FEATURE

BIOLOGICAL INTERPRETATION

● Largest components size → Common targeted regions

● Percolation critical point → Robustness to attacks, e.g. drugs

● Community structure → Quantification of “systemic” degree



Results

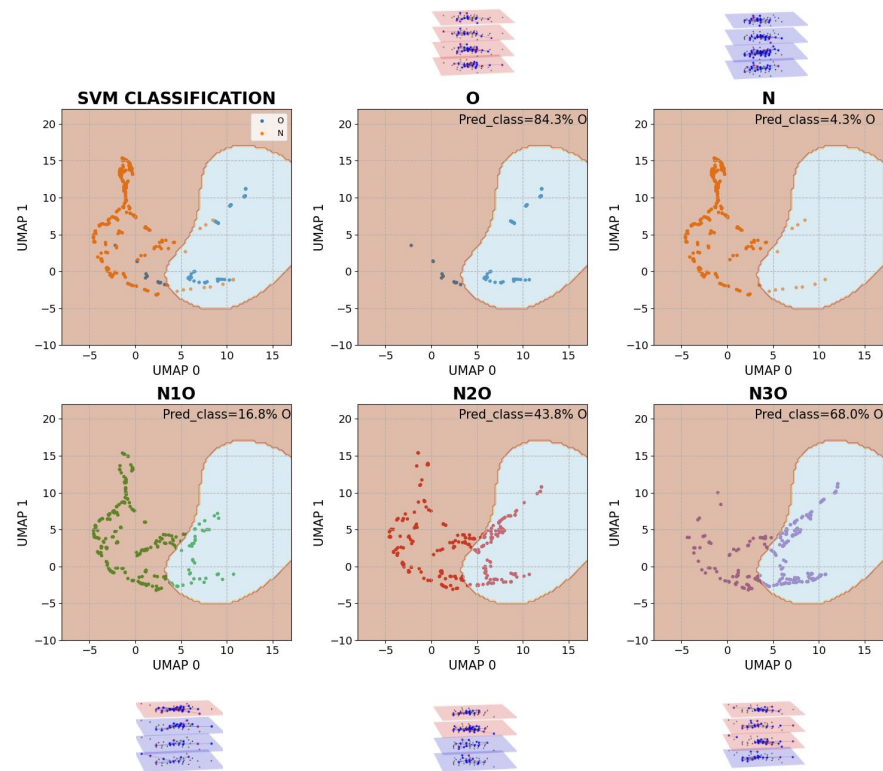
Topological features

Combined features analysis ←

→ UMAP reduced parameter space separation

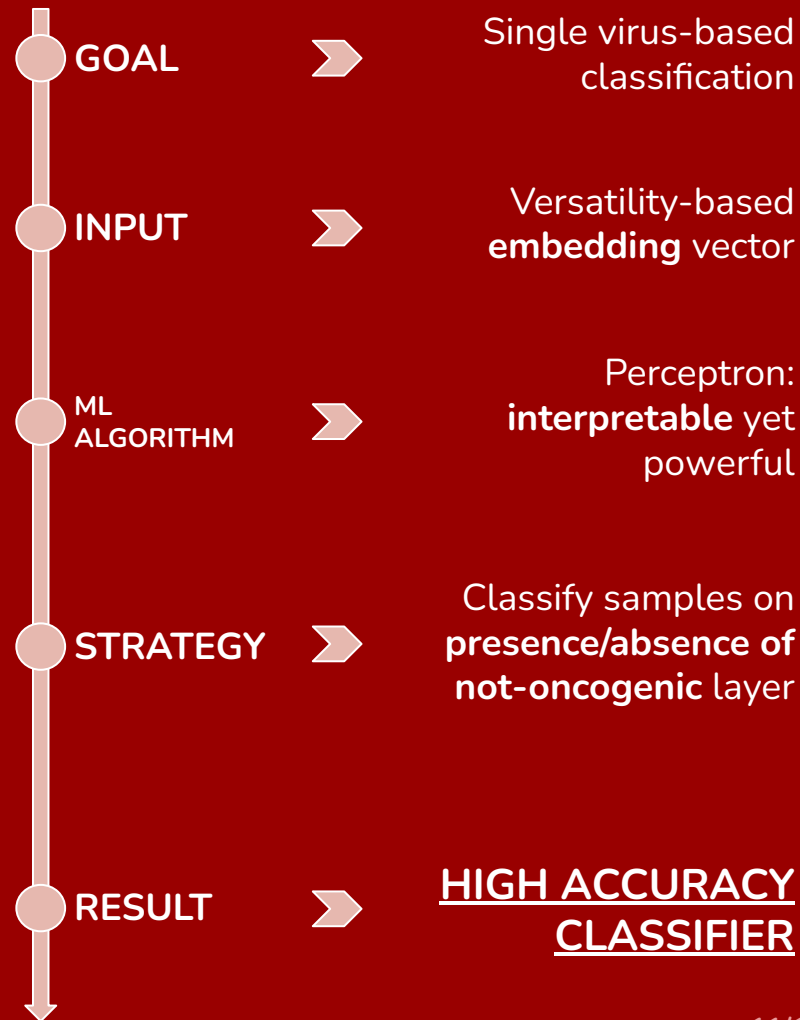
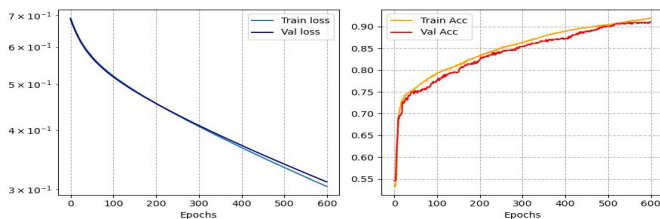
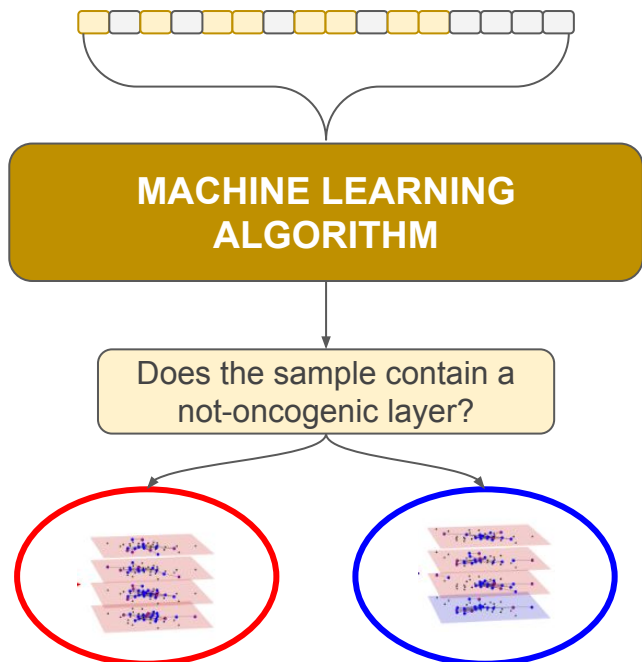
Classification between oncogenic and not-oncogenic regions ←

→ Progressive shift between the two regions



Results

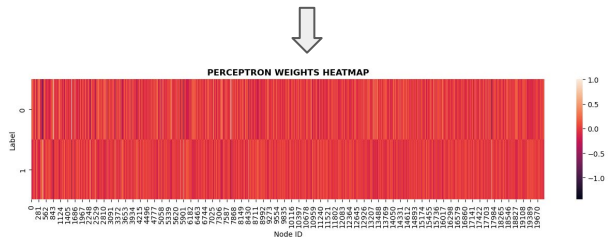
Machine learning



Results

Machine learning

**MACHINE LEARNING
ALGORITHM**



Input features correspond to human proteins

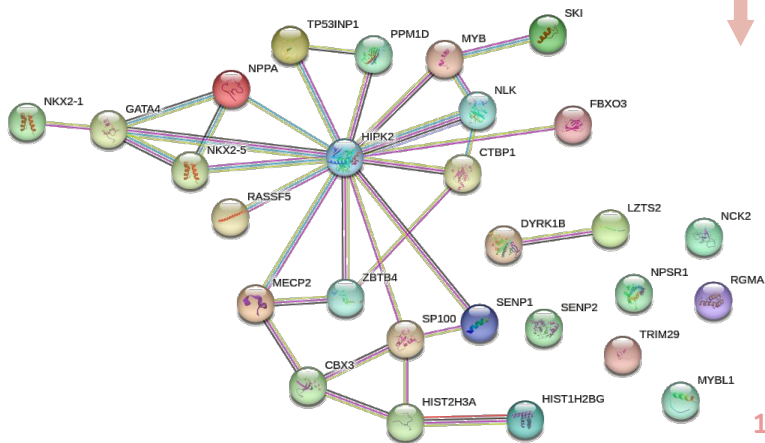
Perceptron weights analysis

Set of proteins potentially connected to some oncogenic mechanisms

**GO Pathway
Enrichment
Analysis**



- gene expression regulation
- chromatin structure



Conclusions

Overview

Twofold approach for oncogenic / not-oncogenic distinction:

- topological feature statistical analysis
- machine learning

Conclusions

- With both methods evidence of relevant distinction between the 2 cases
- Set of proteins with potential connection with oncogenicity in viruses

Perspective

- Classification of early discovered viruses
- Sars-Cov2 analysis ongoing



Thanks for the attention

any questions?