Mining multiple pathogen-host protein interactomes for the detection of shared patterns.

Pieter Moris, Danh Bui-Thi, Pieter Meysman, Kris Laukens

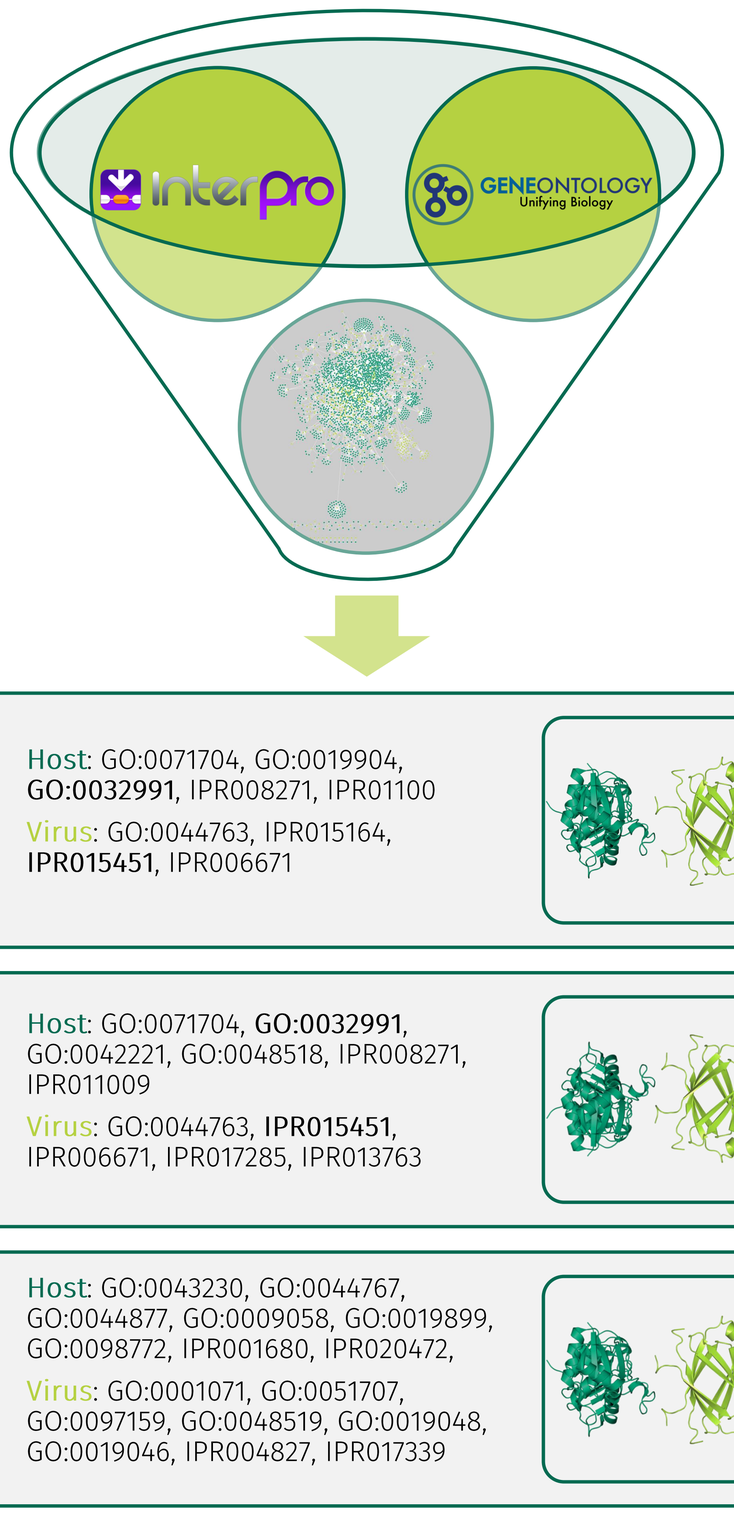
Universiteit Antwerpen

Background

The interplay between intracellular pathogens and their hosts is mediated through intricate molecular interactions. We showcase a case-study on the protein interactomes of *Herpesviridae* and their hosts utilising association rule mining and biological annotation resources. This represents our first steps in the development of a general methodology to analyse pathogen-host networks with the goal of discovering shared characteristics, viral strategies or pathogen-host relationships at the protein level.

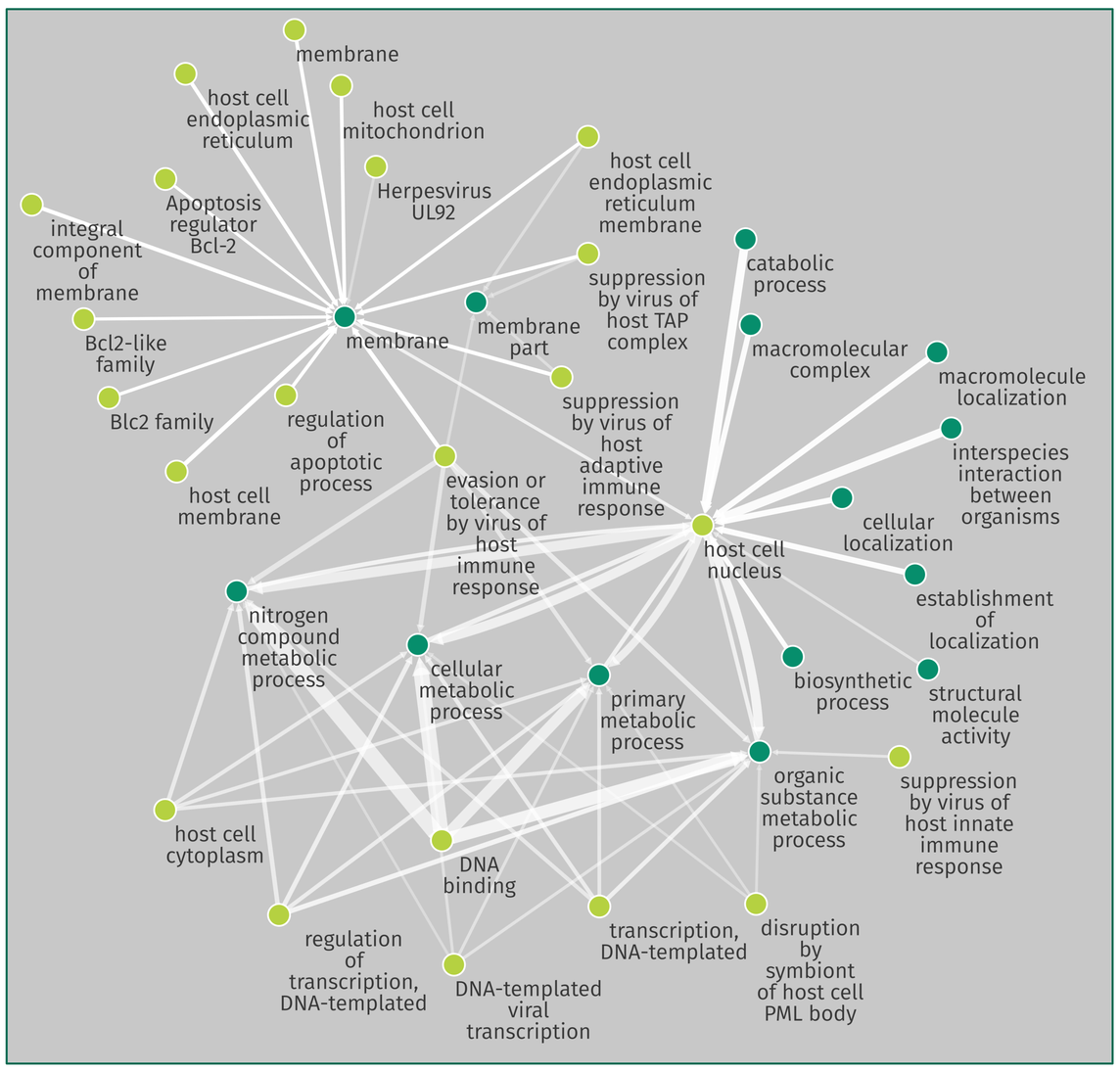
Methodology

- Protein-protein interaction (PPI) networks of 47 herpesviruses and 19 hosts (e.g. *Homo sapiens*, *Mus musculus*, etc.) were collected from HPIDB 2.0, VirHostNet 2.0 and PHISTO.
- Each protein partner was mapped to Gene Ontology annotations and InterPro families, domains and sequence features.
- By considering each PPI as a transaction set with the annotations as its items, frequent item set mining and association rule mining were conducted (Apriori-based approach) in a directional manner (rules going from host to virus or vice versa).



Results

As a summary of the retrieved association rules, all pairwise sub-rules were visualised in a network, after the removal of uninformative generic terms. These rules represent frequent (support > 3%) patterns in the properties of PPIs between *Herpesviridae* and their hosts.



Pairwise sub-rules retrieved by association rule mining.

Light green nodes represent annotations of viral proteins and dark green nodes those of host proteins. Connected nodes signify the existence of an association rule between the annotation terms. The width of the edges is related to the frequency of the sub-rules. More opaque rules have a higher maximum confidence ([0.75; 0.91]).

Discussion and future work

The discovered rules describe frequently occurring annotation patterns that cover some of the broad stages of viral infection:

- The invasion of host cells is visible through interactions with membrane-associated host proteins in the top left cluster.
- Viral shedding and the hijacking of host apoptosis mechanisms is also part of the top left cluster.
- The bottom clusters represent viral replication as shown by the links between viral DNA binding proteins and host metabolic processes.

Association rule mining naturally lends itself to the construction of classification models, thus we plan to apply this methodology to different groups of pathogens to uncover general distinguishing patterns hidden in their PPI networks. In addition, the inclusion of additional (orthogonal) data sources could offer further insight into the patterns that govern the interplay between pathogens and their host.

ADReM group, Department of Mathematics and Computer Science, University of Antwerp, Belgium
Biomedical Informatics Research Center Antwerp (biomina), University of Antwerp & Antwerp University Hospital, Belgium

pieter.moris@uantwerpen.be

github.com/pmoris

@pi_moris