



Contribution ID: 28

Type: Talk

An Interpretable Graph Message Passing Neural Network for predicting disease causing genes in viral infections

Viral infections are complex multisystemic diseases characterized by genetic and molecular alterations in the host. Consequently, Identifying disease-related genes and their underlying molecular mechanisms poses significant challenges. To address these challenges, we have enhanced our methodology by employing Graph Message Passing Neural Networks (MPNNs) that leverage host proteins as input nodes and protein-protein interactions (PPIs) to construct the input graph. Each host protein is characterized by multiple omics describing the effect of a specific viral infection, which we use as node features. We further incorporated functional embeddings from Gene Ontology and positional encodings of each gene in PPI networks into the input features. This enrichment enables more effective classification of unlabeled host proteins based on their relationship with the surrounding neighborhood and associated feature vectors. Building on the established effectiveness of MPNNs in accurately predicting cancer-related genes, we adapted this methodology to study SARS-CoV-2 infections and extended it to less studied viruses. Our approach leverages on a pre-training phase using disease-related genes across all human viral infections, followed by fine-tuning on a specific viral disease, using corresponding omics information as new input features to characterize viral pathogenicity. This methodological evolution, coupled with the implementation of self-supervised learning strategies to effectively handle missing omics information for some of the genes present in the interaction network, positions our model as a versatile tool in the challenge for understanding viral diseases. Specifically for SARS-CoV-2, we integrate transcriptome, proteome, effectome, and virus-host interactome data from A549 cells to identify novel host factors. Our findings have revealed numerous potential host factors that have already been independently validated in publications as potential host factors or antiviral drug targets. To prioritize novel predictions for experimental validation via CRISPR-Cas9 knockout and drug screens, we utilize perturbation-based explanation techniques. This strategy helps us understand the molecular mechanisms that contribute to the classification of a newly discovered host factor and the prioritization of genes for experimental validation. Our model offers a tangible and scalable solution to advance our understanding of poorly characterized viral infections by making use of existing prior knowledge. Most notably, the discovery and experimental validation of novel host factors will facilitate the development and re-purposing of antiviral drugs.

Primary author: FIRMANI, Samuele (Helmholtz Munich)

Co-author: Prof. MARSICO, Annalisa (Helmholtz Munich)

Session Classification: Break + Posters session

Track Classification: Poster