Investigating host dependency factor screening data across hemorrhagic fever and respiratory viruses to identify communalities and differences using machine learning

Alicia Hiemisch[1], Matteo Rizzato[2], Mohadeseh Naseri[1], André Dietz[1], Tuna Toptan[2], Rainer König[1]

1. Jena University Hospital, Institute for Infectious Diseases and Infection Control
2. Technical University of Munich, School of Medicine, Institute of Virology

Introduction

Since emerging in 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic, but it's not the only viral threat to human health. Influenza A virus (IAV) has shown pandemic potential through past outbreaks, and infections from Dengue and West Nile viruses are rising, including Europe (Logiudice, Jacopo et al. 2025). Besides vaccines, host-directed antivirals (HDA) are promising therapeutic options. While CRISPR/Cas9 or RNA interference (RNAi) screens have been employed in numerous studies to identify relevant host factors (HDF), achieving robust and consistent identification poses a challenge (Edinger et al. 2014; Wagoner et al. 2022). These studies are lacking consistency, making it difficult to interpret the results in a broad-spectrum context. In the present work, host dependency factor screening data from 76 published research articles were analyzed in an integrated approach. A boosted random forest model was employed to detect conserved patterns within each viral group, revealing reveal virus-specific and cross-viral pathways.

Methods/Results

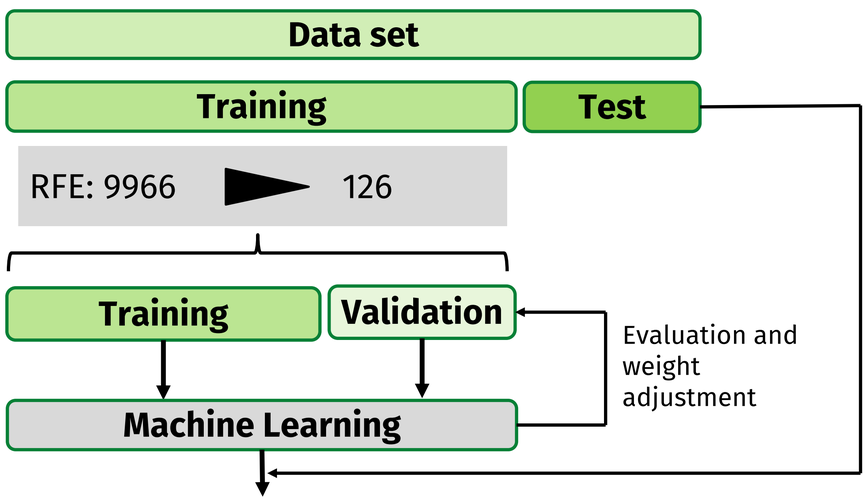
Literature research and feature table construction

Causative agent	#data sets	Viral category
SARS corona virus 2	21	SARS2
Influenza A virus	16	IAV
Denque virus	14	Viral hemorrhagic fever (VHF)
Zika virus	11	
Ebola virus	5	
West Nile virus	4	
Yellow fever virus	4	
Rift valley fever virus	3	

GSEA

Screen	HDF1	HDF2	HDF3	Gene Set 1	Gene Set 2	Etc.	Label
Screen A	1	0	0	0	1	...	SARS2
Screen B	1	1	1	1	0	...	IAV
Screen C	0	1	1	0	1	...	VHF
Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.

Boosted random forest framework effectively classifies IAV, SARS2 and VHF screens



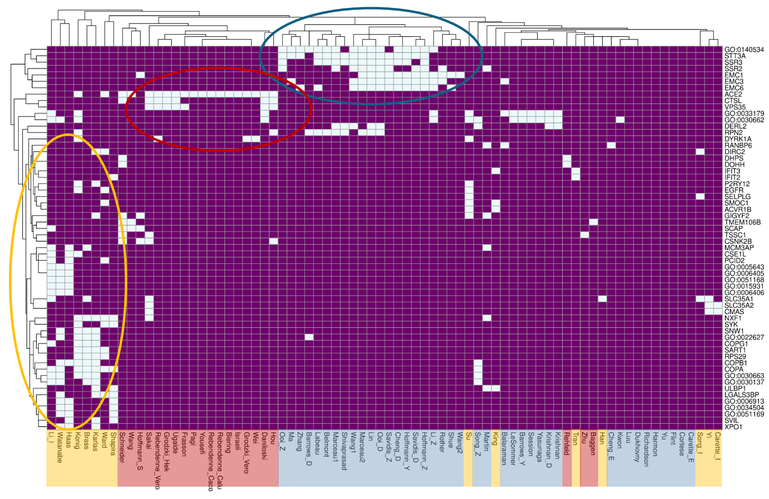
Reference

Prediction	IAV	SARS2	VHF
IAV	8	0	1
SARS2	1	19	0
VHF	6	1	39

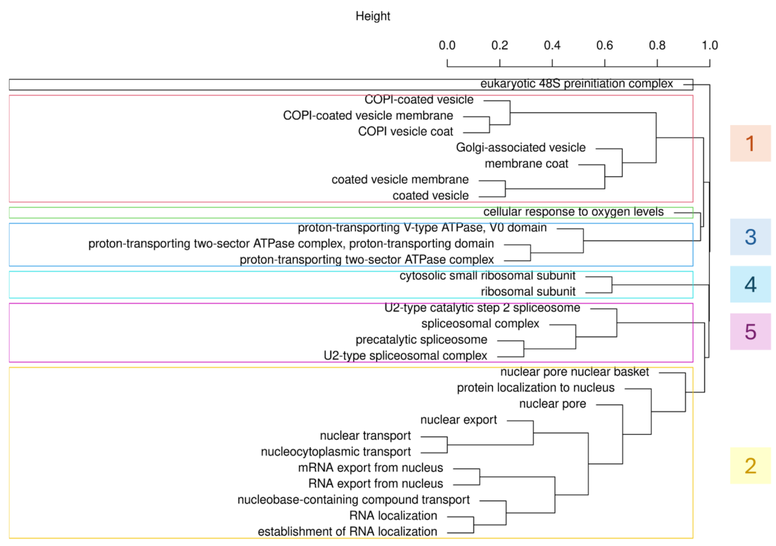
The model achieved a balanced accuracy of 0.88, with class-specific accuracies of 0.76 for IAV, 0.97 for SARS2, and 0.89 for VHF.

Weighted hierarchical clustering of high-impact features reveals virus-specific similarities

The SARS-CoV-2 cluster is primarily characterized by the presence of ACE2. The VHF cluster distincts by its overlap in genes involved in the ER protein-containing complex and the corresponding enriched gene set. Interestingly, the IAV cluster is predominantly defined by diverse gene sets.

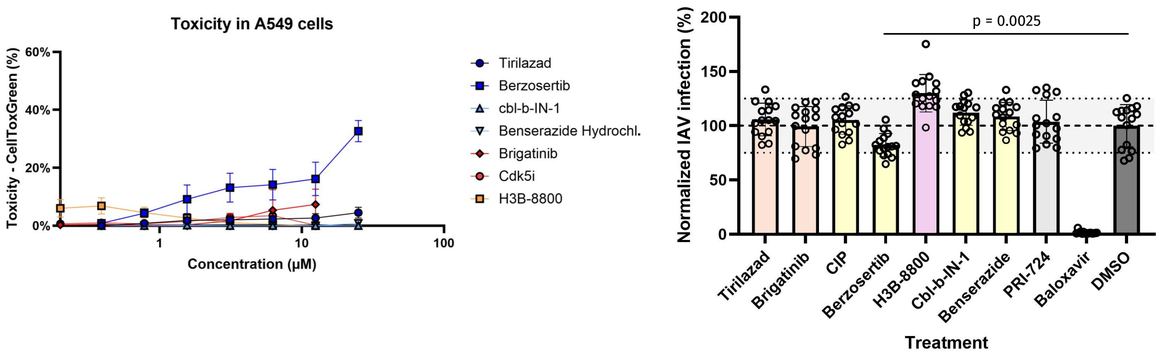


IAV-related gene sets cluster by transport, translation, and splicing processes



Hierarchical clustering of gene sets (Jaccard similarity) revealed seven functional clusters spanning nuclear/vesicle transport (COPI-coated), ATP synthesis/pH regulation, RNA splicing, and ribosomal processes. Drugs were selected targeting proteins in these clusters and prioritized via network-based drug repurposing (Cheng et al., 2018).

Network-predicted berzosertib reduced viral load in IAV-infected A549 cells



Discussion

Host dependency factor screening data show high variability at the gene level, while consistency improves when focusing on significantly enriched gene sets. A boosted random forest model distinguished VHF viruses, SARS-CoV-2, and IAV with 88% balanced accuracy. Feature-based clustering revealed an IAV-specific group characterized by COPI-coated vesicle transport, nuclear trafficking, proton transport, ribosomal function, and splicing. Recovery of known HDFs (e.g., ACE2, STT3A, XPO1) supports the validity of the approach and highlights virus-specific host dependencies. Finally, a network-based drug repurposing pipeline prioritized candidate compounds, and berzosertib significantly reduced viral load in IAV-infected A549 cells, demonstrating the translational potential of this strategy.

References

- Cheng et al.; Nat Commun. 2018; 9(1):2691.
- Edinger et al.; J. Gen. Virol. 2014; 95: 26.
- Logiudice, Jacopo et al.; Pathogens 2025; 14,(1)63.
- Wagoner et al., Microbiol Spectr. 2022; 10:e0333122

SCAN ME

Funding: The APPEAL project receives funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101137311. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union