APPEAL
(Antivirus Pandemic Preparedness EuropeAn pLatform)
Data-Driven Identification of Host Targets for Broad-Spectrum Antivirals

Andre Dietz[1], Mohadeseh Naseri[1], Mathieu Mateo[2], Seyoum Ayehunie[3], Holger Erfle[4], Tuna Toptan Grabmair[5], Caroline Goujon[6], Rainer König[1]

[1] Jena University Hospital, Institute for Infectious Diseases and Infection Control, Jena, Germany; [2] Institut Pasteur, Lyon, France; [3] Mattek Europe, Bratislava, Slovakia; [4] Ruprecht Karls University, Heidelberg, Germany; [5] Technical University of Munich, Munich, Germany; [6] Institut de Recherche en Infectiologie de Montpellier (IRIM), CNRS, France

Introduction

Viruses critically depend on host cell proteins, so-called host dependency factors (HDF), for successful infection and replication. While traditional antivirals target viral components directly, this approach is vulnerable to resistance through rapid viral mutation. In contrast, targeting HDF offers the potential for broad-spectrum antiviral therapies with a reduced risk of resistance development (1).

However, identifying reliable, broadly acting HDF is a major challenge. Although numerous genome-wide knockout (KO) and knockdown (KD) screens have been published for various high-risk viruses—including SARS-CoV-2, SARS-CoV-1, MERS, Influenza A, Dengue, Zika, Rift Valley Fever Virus, and Ebola—these datasets are highly heterogeneous. Differences in cell lines, viral strains, infection protocols, and readout strategies result in limited overlap between identified HDF, even for the same virus (2).

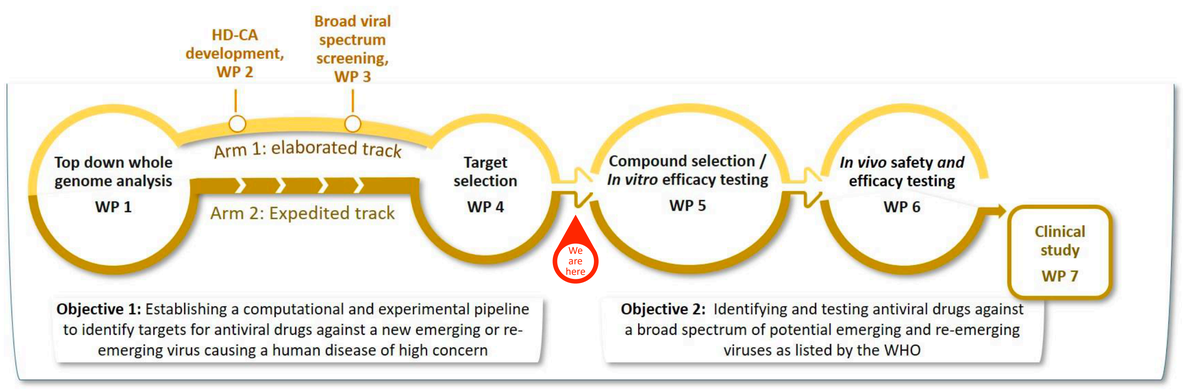
To address this, we systematically analyzed 90 published KO/KD screens across the mentioned eight viral species. The central challenge was to integrate these heterogeneous datasets in a way that preserves meaningful biological signal while filtering out noise. To that end, we implemented and compared several data aggregation strategies and evaluated their robustness using cross-validation. This approach allowed us to identify a consensus list of 402 candidate HDF with potential pan-viral relevance.

Within the APPEAL project, a fast-track pipeline focuses on ~20–30 of these candidates that are already targeted by approved drugs. These will be prioritized for validation, with the ultimate goal of progressing to clinical evaluation as broad-spectrum antiviral treatments.

1 The APPEAL Project: A Dual-Track Strategy for Pandemic Preparedness

The APPEAL project combines computational and experimental approaches to develop host-directed antiviral therapies (HDAs) that are effective against a broad range of emerging and re-emerging viruses prioritized by the WHO (3).

The project is structured into seven work packages (WP1–WP7), progressing from in silico genome analysis and host-target identification to compound screening, preclinical validation, and ultimately clinical evaluation.



The APPEAL project follows two complementary tracks:

- Arm 1 – Elaborated Track: A comprehensive and stepwise process that ensures robust validation of promising antiviral targets and compounds.
- Arm 2 – Expedited Track: A fast-track route focusing on drug repurposing to rapidly identify and deliver effective treatments in response to urgent pandemic situations.

2 KO/KD Screens Reveal Major Inconsistencies

A total of 90 KO/KD screens targeting HDF were compiled from published studies across eight viruses, including SARS-CoV-2, SARS-CoV-1, Influenza A, Dengue, Zika, MERS, Ebola, and Rift Valley fever virus. The datasets varied widely in completeness and format: some provided only hit lists, others included full gene-level data, with or without p-values or statistical rankings. Table 1 summarizes the retrieved screens, including the number of datasets per virus, datasets without p-values, gene rankings, or defined screening sets.

Table 1: Distribution of KO/KD screening data for the explored viruses.

Virus	Total number of screens	Number of screens without p-value	Number of screens without ranking	Number of screens without defined set of screened genes
SARS-CoV-2	29	1	1	0
IAV	21	13	7	4
Dengue	14	7	3	0
Zika	11	3	2	1
MERS/SARS-CoV	6	0	0	0
EBOV	6	1	1	1
RVFV	3	3	0	0
Sum	90	28	14	6

Funded by the European Union

Horizon Europe research and innovation program (Grant Agreement No. 101137311)
www.project-appeal.eu

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 granting authority can be held responsible for them.

3 Combining Heterogeneous Screens: Six Approaches to Unify the Data

To overcome inconsistencies across the 90 KO/KD screens and identify robust, pan-viral HDF, six complementary aggregation strategies were tested:

1. **Voting by Frequency**
   Genes selected as HDFs in more than seven screens were retained.
2. **MAIC Regression**
   A published tool for linear regression (MAIC) (4) was applied to integrate hits across studies.
3. **P-Value Aggregation Score**
   Each gene was scored by summing the –log(p-value) across all screens reporting it.
4. **Best GSEA Enrichment P-Value x P-Value Aggregation Score**
   For each screen, Gene Set Enrichment Analysis (GSEA) was performed using g:Profiler. Genes were weighted by the significance of their first gene set hit, then multiplied by their p-value aggregation score (from approach 3).
5. **Total GSEA Enrichment P-Value x P-Value Aggregation Score**
   Similar to (4), but enrichment scores from all gene sets were summed per gene before multiplication with the p-value aggregation score.
6. **GSEA + Frequency Filter**
   Gene sets occurring in ≥ 8 screens were retained. Genes from these sets were selected if also reported as HDFs in > 4 screens.

4 Evaluating Robustness: Cross-Validation of Aggregation Strategies

To assess the performance of the six approaches for generating ranked HDF lists, a cross-validation strategy was applied:

- In each run, the 90 screens were randomly split into 80 training and 10 test screens.
- A ranked list of 1,000 candidate HDFs was generated from the training set.
- The overlap with each test set was evaluated using a chi-square test.
- This process was repeated 100 times per method, yielding 1,000 p-values per approach.
- The median p-value served as the performance metric for comparison (Table 2).

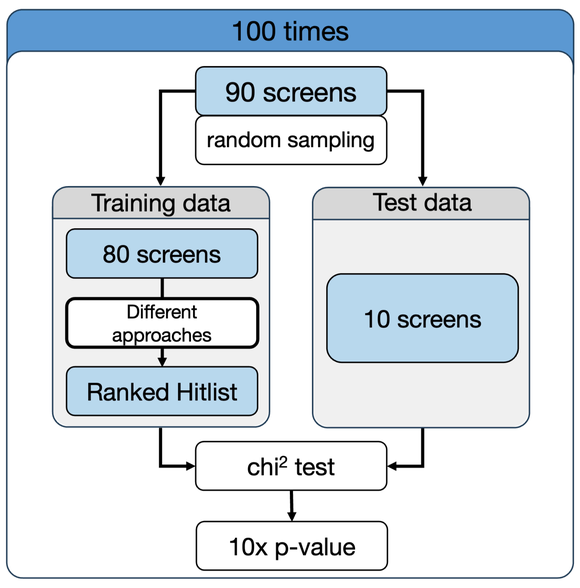


Table 2: Median of 1000 p-values for different aggregation approaches.

Aggregation approach	Median p-value of top 1000 HDF vs test data
1. HDF in > 7 screens	$2x10^{-5}$
2. MAIC	0.015
3. P-Value Aggregation Score	$1x10^{-5}$
4. Best GSEA Enrichment P-Value x P-Value Aggregation Score	$2x10^{-5}$
5. Total GSEA Enrichment P-Value x P-Value Aggregation Score	$3x10^{-5}$
6. GSEA + Frequency Filter	$1x10^{-5}$

Aggregation of the Final HDF List

To generate a final list, the top 1,000 genes from the four best-performing approaches (1, 3, 4, and 6) were combined. Genes found in ≥ 3 of the 4 lists were retained, resulting in 402 candidate HDF sorted by products of their ranks.

A permutation test was used to evaluate significance, identifying 84 high-confidence HDF (FDR < 0.05). The remaining 318 genes were considered lower-priority but included in downstream analyses.

To assess biological relevance and druggability, HDF were annotated with data on gene essentiality, direct drug targets, complex or pathway involvement, and crystal structure availability (Table 3).

Table 3: Overview of high- and lower-priority HDFs, including counts for gene essentiality, druggability, structural data, and functional enrichment.

	Number	HDF essential for cell survival	HDF with direct targeting drugs	HDF with known crystal structure	HDF with known crystal structure but w/o direct targeting drug
HDF with FDR < 0.05	84	5	29	38	21
HDF with FDR > 0.05	318	52	100	135	69
Sum	402	57	129	173	90

References

(1) Kaufmann, S. H. E., Dorhoi, A., Hotchkiss, R. S. & Bartenschlager, R. Host-directed therapies for bacterial and viral infections. Nat. Rev. Drug Discov. 17, 35–56 (2018).
(2) Kelch, M. A. et al. Machine learning on large scale perturbation screens for SARS-CoV-2 host factors identifies β-catenin/CBP inhibitor PRI-724 as a potent antiviral. Front. Microbiol. 14, 1193320 (2023).
(3) WHO. Prioritizing Diseases for Research and Development in Emergency Contexts. https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts (2025).
(4) Signorovitch, J. E. et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res. 15, 940–947 (2012).