

Pathogens Preferentially Target High-Centrality Host Proteins: A Network-Topology and Functional Characterisation of the Human Pathogen–Host Interactome

LitLit AI

LitLit AI, contact@litlit.ai

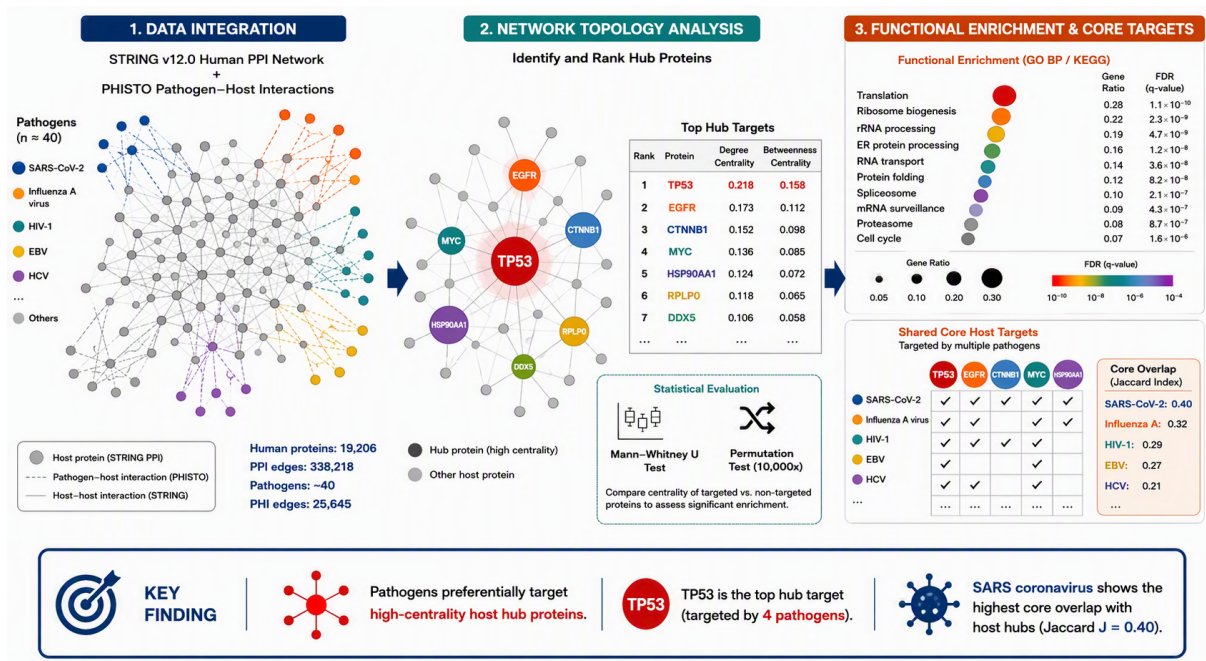


Figure 1: **Graphical abstract.** Overview of the analysis. The STRING v12.0 human PPI backbone and the PHISTO/Ahmed et al. pathogen–host interaction catalogue are integrated; every host protein is annotated with degree and betweenness centrality and with pathogen-target status; Mann–Whitney and degree-matched permutation tests establish that pathogen targets are enriched among network hubs; GO/KEGG enrichment identifies translation, ribosome, ER processing, and RNA transport as the dominant targeted functions; a multi-pathogen “core” target set and per-pathogen Jaccard overlap reveal that coronaviruses draw most heavily on the shared core and that TP53 is the top hub target.

Abstract

Pathogens do not perturb single host factors; they rewire subgraphs of the human protein interaction network, and the topological position of the host proteins they engage governs the efficiency of that rewiring. Whether this hub-targeting logic holds across a broad, phylogenetically diverse panel of pathogens, and which host proteins constitute the shared contact points of convergent exploitation, remain incompletely resolved. Here we integrated the STRING v12.0 human

protein–protein interaction (PPI) backbone (16,059 nodes, 234,813 edges; confidence ≥ 0.7) with PHISTO, a curated catalogue of $\sim 5,500$ pathogen–host interactions spanning 156 pathogens. We annotated every host protein with degree and betweenness centrality, tested whether pathogen targets are enriched among hubs using both a Mann–Whitney U test and a degree-matched permutation null model (1,000 permutations, seed 42), performed Gene Ontology and KEGG functional enrichment, defined a multi-pathogen “core” target set, and quantified per-pathogen overlap by the Jaccard coefficient. Pathogen-targeted proteins (2,453; 15.3% of nodes) had significantly higher degree (median 28 vs 12, $p = 1.7 \times 10^{-146}$, Cliff’s $\delta = 0.326$) and betweenness ($p = 5.9 \times 10^{-126}$) than non-targets, and remained over-represented among the top-5% hubs even under the degree-matched null ($p = 0.0010$), ruling out degree as the sole driver. Functional enrichment (621 terms, FDR < 0.05) converged on host translation and the ribosome (top GO Biological Process: Translation, OR 4.53; top KEGG: Ribosome, OR 5.28), proteostasis, and the secretory pathway. The shared core (≥ 2 pathogens) comprised 933 proteins, narrowing to 30 at ≥ 5 ; DLG1 was the most broadly targeted host protein (25 pathogens). The top hub target was TP53 (betweenness 0.0428, degree 766, four pathogens), followed by CTNNB1, EGFR, and ACTB. SARS coronavirus showed the highest Jaccard overlap with the core set ($J = 0.402$), followed by SARS-CoV-2 ($J = 0.393$). These results demonstrate that pathogens convergently target the most central hubs of the human interactome—master regulators of signalling, the translational machinery, and cell architecture—and identify a shared core of universally exploited host contact points with implications for host-directed, broad-spectrum therapeutic strategies. The enrichment is highly significant and robust to a degree-matched null, though small in effect size (Cliff’s $\delta \approx 0.30$), indicating that pathogens draw preferentially—but not exclusively—from the hub stratum.

1 Introduction

Infectious diseases remain among the leading causes of human morbidity and mortality worldwide, and the outcome of any host–pathogen encounter is governed less by the pathogen alone than by the molecular interface at which pathogen-encoded proteins engage the host cell. Over the past two decades, high-throughput yeast two-hybrid, affinity-purification mass spectrometry, and protein-fragment complementation assays have made it possible to map protein–protein interactions (PPIs) at near-genome scale, yielding comprehensive maps of both the human interactome [1–4] and the physical contacts that diverse pathogens establish with human proteins [5–9]. These maps have reframed infectious disease as a problem in network biology [10, 11]: a pathogen does not perturb a single host factor but rewires a subgraph of the host interactome, and the topological position of the host proteins it engages determines the breadth and efficiency of that rewiring [12, 13].

A central organising principle of biological networks is the centrality–lethality relationship: the most highly connected nodes—hubs—tend to correspond to essential genes, and their disruption is disproportionately lethal [14–17]. Extending this logic to host–pathogen systems, several studies have reported that viral and bacterial proteins preferentially interact with host hubs, a strategy that maximises the pathogen’s reach over host regulatory and metabolic circuits while minimising the number of contacts it must establish [6, 12, 13, 18]. Hubs are attractive targets for a pathogen for the same reason they are dangerous to delete: a single interaction with a high-degree node grants access to a large fraction of the network, and high-betweenness “bridge” nodes confer control over information flow between otherwise distant modules [11, 19]. Yet the generality of this pattern across a broad taxonomic spectrum of pathogens—encompassing viruses, bacteria, and protozoa—and the functional identity of the shared host hubs that disparate pathogens converge upon, remain incompletely characterised. Earlier surveys were often restricted to a single pathogen class, relied on now-outdated interactome maps, or lacked a

degree-controlled null model that can separate genuine hub-preference from the trivial fact that any reasonably central protein set will, on average, be more connected than the network at large. The resource landscape has since matured. STRING v12.0 [20] now provides a quality-scored, continuously updated human PPI backbone that integrates experimental, curated, and text-mined evidence; and PHISTO—the curated pathogen–host interaction catalogue of Ahmed et al. [5]—consolidates experimentally validated pathogen–host interactions for hundreds of pathogens into a single, manually curated catalogue. Together these two resources enable, for the first time, a unified, network-analytic interrogation of how a phylogenetically diverse panel of pathogens targets the modern human interactome.

Here we integrate STRING v12.0 and PHISTO to ask three questions. First, are pathogen-targeted host proteins enriched among the high-degree and high-betweenness hubs of the human PPI network, and does this enrichment survive a degree-matched null model that controls for the targets’ overall degree distribution? Second, what host functions and pathways do pathogen targets converge upon, and what does that functional convergence reveal about the shared virulence strategies of phylogenetically distant pathogens? Third, do individual pathogens differ in the degree to which their target repertoires overlap with a shared “core” set of multi-pathogen host targets, and which host proteins constitute the most universally exploited contact points? To answer these questions we constructed the high-confidence human PPI backbone, annotated every protein with degree and betweenness centrality and with its pathogen-target status, applied distributional (Mann–Whitney) and degree-controlled (permutation) enrichment tests, performed GO and KEGG functional enrichment, defined a multi-pathogen core target set, and quantified per-pathogen overlap by the Jaccard coefficient. Our findings indicate that pathogens preferentially target the most central hubs of the human interactome—with TP53, EGFR, β -catenin, and the translational machinery heading the list—and that coronaviruses, in particular SARS coronavirus, draw their host target repertoires most heavily from the shared core, pointing to a convergent, network-centred logic of pathogen exploitation.

2 Results

2.1 The human PPI backbone is a sparse, well-connected network

After filtering the STRING v12.0 human interaction table to interactions with combined confidence ≥ 0.7 , mapping identifiers to HGNC gene symbols, and removing self-loops and duplicate undirected pairs, the resulting backbone comprised 16,059 nodes connected by 234,813 unique undirected edges (Figure 1; Table 1). The network was sparse (density = 1.82×10^{-3}) yet well connected: 127 connected components were present, but the giant connected component alone contained 15,741 nodes (98.02% of all proteins), indicating that the high-confidence human interactome is essentially a single macroscopic network with a long tail of small isolated subgraphs. The average clustering coefficient was 0.377, substantially higher than expected for an Erdős–Rényi random graph of the same size and density, consistent with the modular, small-world, scale-free organisation characteristic of biological PPI networks [14, 21]. Degree was heavy-tailed (minimum 1, maximum 766, mean 29.2, median 13.0, standard deviation 45.5), and betweenness centrality was similarly skewed (maximum 0.0428, mean 1.73×10^{-4} , median 2.12×10^{-5}).

Of the 2,649 unique host proteins recorded in PHISTO, 2,453 (92.6%) were present in the STRING backbone, confirming near-complete coverage of the curated pathogen–host interactome within the high-confidence human PPI network. These 2,453 pathogen-targeted proteins therefore constitute 15.27% of all network nodes, with the remaining 13,606 nodes serving as non-target controls for all subsequent enrichment tests.

Table 1: Network statistics of the STRING v12.0 human PPI backbone (confidence ≥ 0.7) and pathogen-target annotation.

Property	Value
Nodes	16,059
Edges (unique undirected)	234,813
Connected components	127
Giant component size (% of nodes)	15,741 (98.02%)
Network density	1.82×10^{-3}
Average clustering coefficient	0.377
Degree: min / median / mean / max	1 / 13.0 / 29.24 / 766
Betweenness: min / median / mean / max ($\times 10^{-3}$)	0 / 0.0212 / 0.173 / 42.84
Pathogen-targeted nodes in STRING	2,453 (15.27%)
Non-target nodes	13,606
PHISTO host proteins covered by STRING	2,453 / 2,649 (92.6%)

2.2 Pathogen targets are enriched among high-degree and high-betweenness hubs

Pathogen-targeted host proteins showed significantly higher degree and betweenness centrality than non-target proteins. The median degree of pathogen targets was 28, more than double the median of 12 for non-targets (Mann–Whitney $U = 2.21 \times 10^7$, one-sided $p = 1.74 \times 10^{-146}$; Cliff’s $\delta = 0.326$, small effect; Figure 2). The same pattern held for betweenness centrality: pathogen targets had a median betweenness of 6.72×10^{-5} versus 1.62×10^{-5} for non-targets ($U = 2.17 \times 10^7$, $p = 5.92 \times 10^{-126}$; $\delta = 0.301$, small effect; Figure 3). These highly significant differences indicate that pathogens do not interact with host proteins at random but preferentially engage nodes positioned at the densely connected, topologically central core of the human interactome [6, 12, 22].

Because the Mann–Whitney test compares whole distributions, the enrichment could in principle be driven entirely by the fact that pathogen targets have higher degree on average, rather than by clustering at the very apex of the centrality hierarchy. The degree-matched permutation null model controls for this directly: by sampling, within each degree decile, an equal number of non-target nodes with the same degree distribution as the targets, it isolates the residual tendency of targets to occupy the top-5% hub stratum. Even under this stringent, degree-controlled null, pathogen targets remained significantly over-represented among the top-5% degree hubs (observed target hub proportion 0.135 versus a non-target baseline of 0.035 and a degree-matched permutation mean of 0.101; observed/null ratio 1.35; empirical $p = 0.0010$, 1,000 permutations, seed 42; at the resolution floor of 1,000 permutations the smallest reportable p is ≈ 0.001) and among the top-5% betweenness hubs (observed 0.104 vs. non-target 0.040 and null mean 0.088; ratio 1.19; $p = 0.0010$; Figure 4). The residual enrichment within comparable degree strata demonstrates that the pathogen-target bias is not merely a consequence of targets having higher degree overall; pathogens preferentially engage the very highest-degree and highest-betweenness nodes even after accounting for degree.

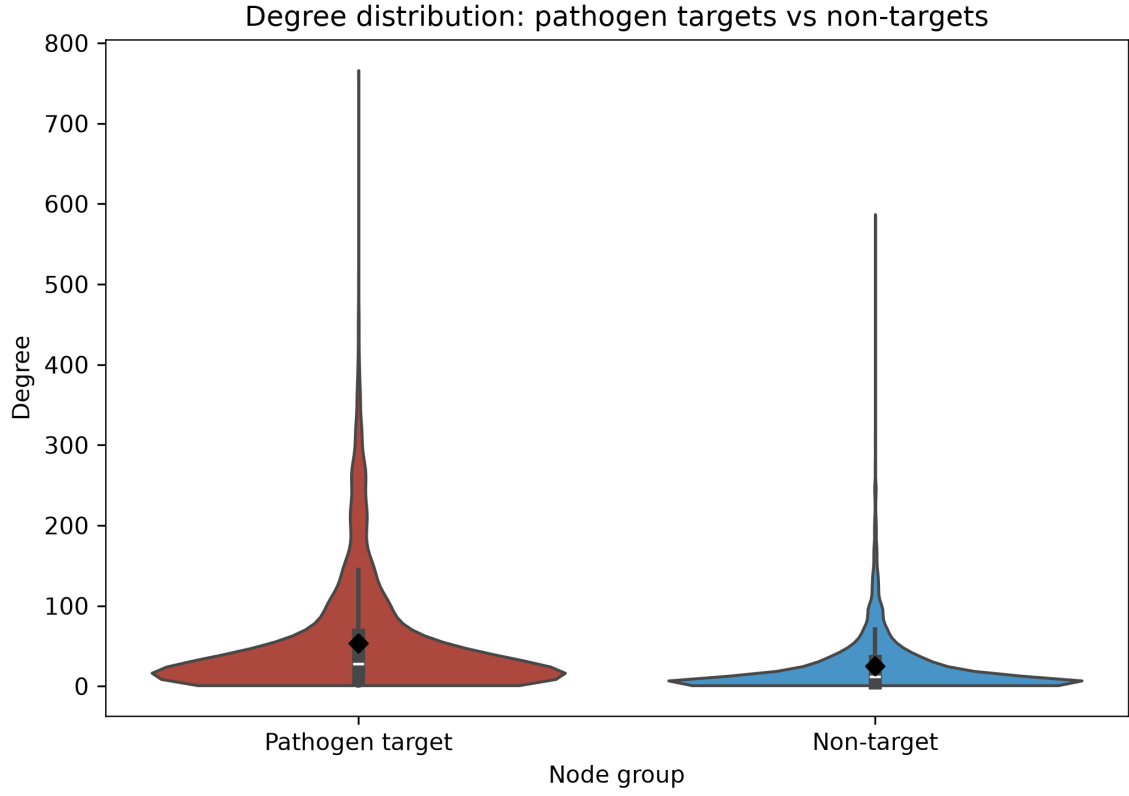


Figure 2: **Degree distributions of pathogen targets versus non-target host proteins.** Pathogen-targeted proteins (orange) are shifted toward higher degree compared with non-target proteins (blue), consistent with preferential targeting of network hubs. Distributions are shown on a logarithmic x -axis. The difference is highly significant (Mann–Whitney $U = 2.21 \times 10^7$, $p = 1.74 \times 10^{-146}$, Cliff's $\delta = 0.326$).

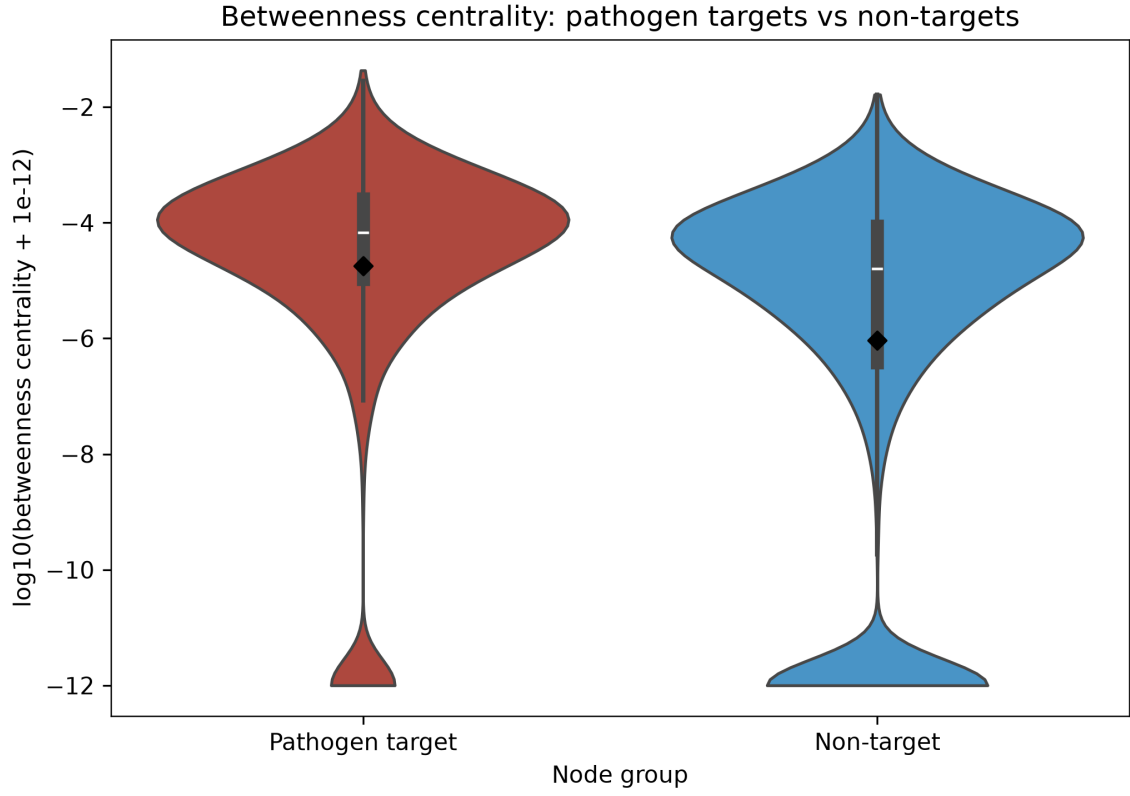


Figure 3: **Betweenness centrality distributions of pathogen targets versus non-target host proteins.** Pathogen-targeted proteins show higher betweenness centrality than non-targets, indicating they occupy more bridge-like, information-flow-critical positions in the network ($U = 2.17 \times 10^7$, $p = 5.92 \times 10^{-126}$, Cliff's $\delta = 0.301$). Betweenness was estimated by approximate sampling ($k = 1,000$ source nodes, seed 42).

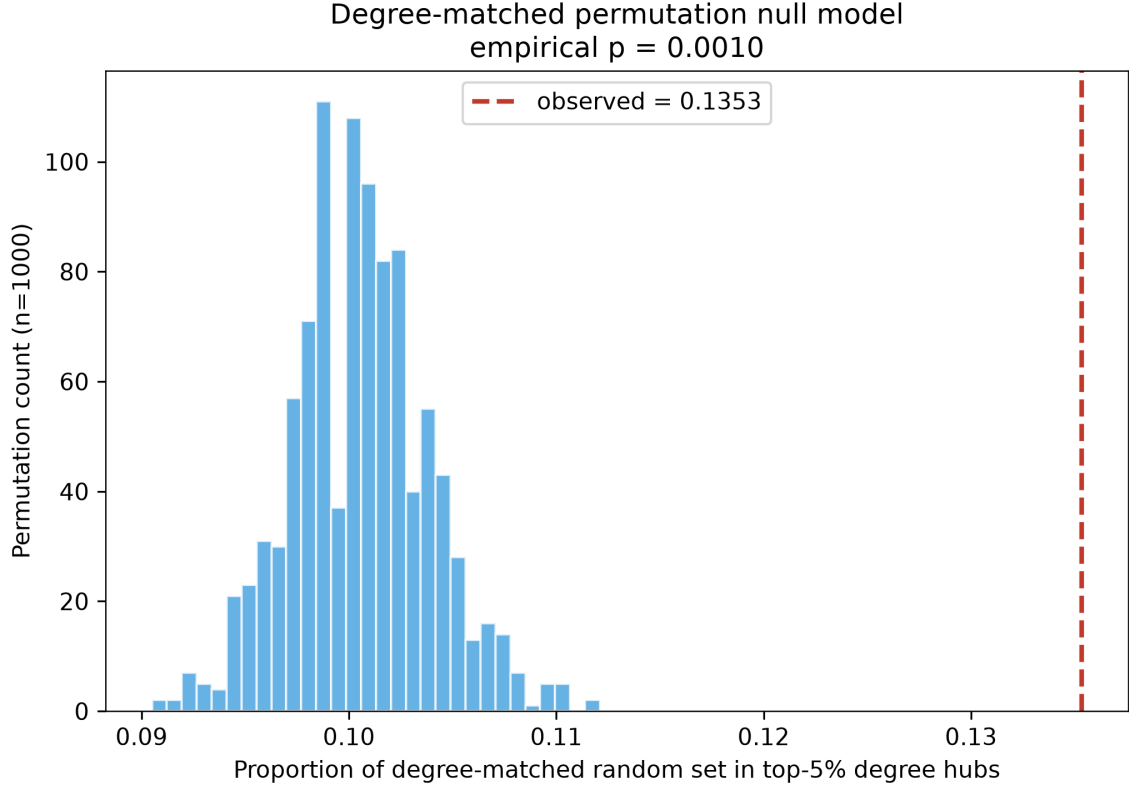


Figure 4: **Degree-matched permutation null model.** The observed proportion of pathogen targets falling in the top-5% hubs (red) exceeds the permutation distribution of degree-matched random sets (histograms) for both degree (left) and betweenness (right) hubs, even though the random sets are constructed to have the same degree distribution as the targets. Empirical $p = 0.0010$ for both (1,000 permutations, seed 42), confirming that pathogen targets are enriched at the apex of the centrality hierarchy independently of their degree.

2.3 Pathogen targets converge on host translation, proteostasis, and secretory pathways

Functional over-representation analysis of the 2,453 pathogen-target genes against the 16,059-gene STRING background yielded 621 significant gene sets at $FDR < 0.05$ (340 Biological Process, 126 Cellular Component, 47 Molecular Function, and 108 KEGG pathway terms). The dominant biological signal was unambiguous and coherent: the single strongest Gene Ontology Biological Process term was *Translation* (GO:0006412; odds ratio [OR] = 4.53, $k = 160$, $FDR = 1.21 \times 10^{-73}$), and the strongest KEGG pathway was *Ribosome* (OR = 5.28, $k = 108$, $FDR = 1.93 \times 10^{-61}$), with cytoplasmic and mitochondrial translation, ribosome biogenesis, and peptide/gene-expression biosynthetic processes all ranking among the top BP hits (Figure 5, Figure 6). The top Molecular Function term was *RNA binding* (GO:0003723; OR = 2.57, $k = 523$, $FDR = 1.32 \times 10^{-108}$), accompanied by cadherin binding, mRNA binding, and ubiquitin-ligase-binding activities (Figure 7). At the Cellular Component level, targets were enriched in intracellular non-membrane-bounded organelles, the nucleolus, mitochondrial and endoplasmic reticulum (ER) membranes, and the cytosolic large ribosomal subunit.

A second, proteostasis-and-secretory-pathway cluster was equally clear: *Protein processing in endoplasmic reticulum* (KEGG; OR = 3.02), intracellular protein transport, Golgi vesicle transport, *N*-glycan biosynthesis, the spliceosome, and RNA transport were all significantly over-represented. Infection-relevant KEGG pathways—*Coronavirus disease*, *Bacterial invasion of ep-*

117 *ithelial cells*, and *Salmonella infection*—were also enriched, as were neurodegeneration-associated
 118 pathways (Alzheimer disease, amyotrophic lateral sclerosis), all of which share NF- κ B, PI3K–
 119 AKT, cytoskeletal, and proteasome components. Taken together, the enrichment landscape
 120 shows that pathogens converge on a shared functional core of host cell biology: the translational
 121 machinery they must commandeer to produce pathogen proteins, the ER folding, glycosyla-
 122 tion, and vesicular-trafficking infrastructure required to mature and traffic those proteins, and
 123 the innate-immune and cytoskeletal signalling modules they must rewire to establish infection
 124 [23, 24].

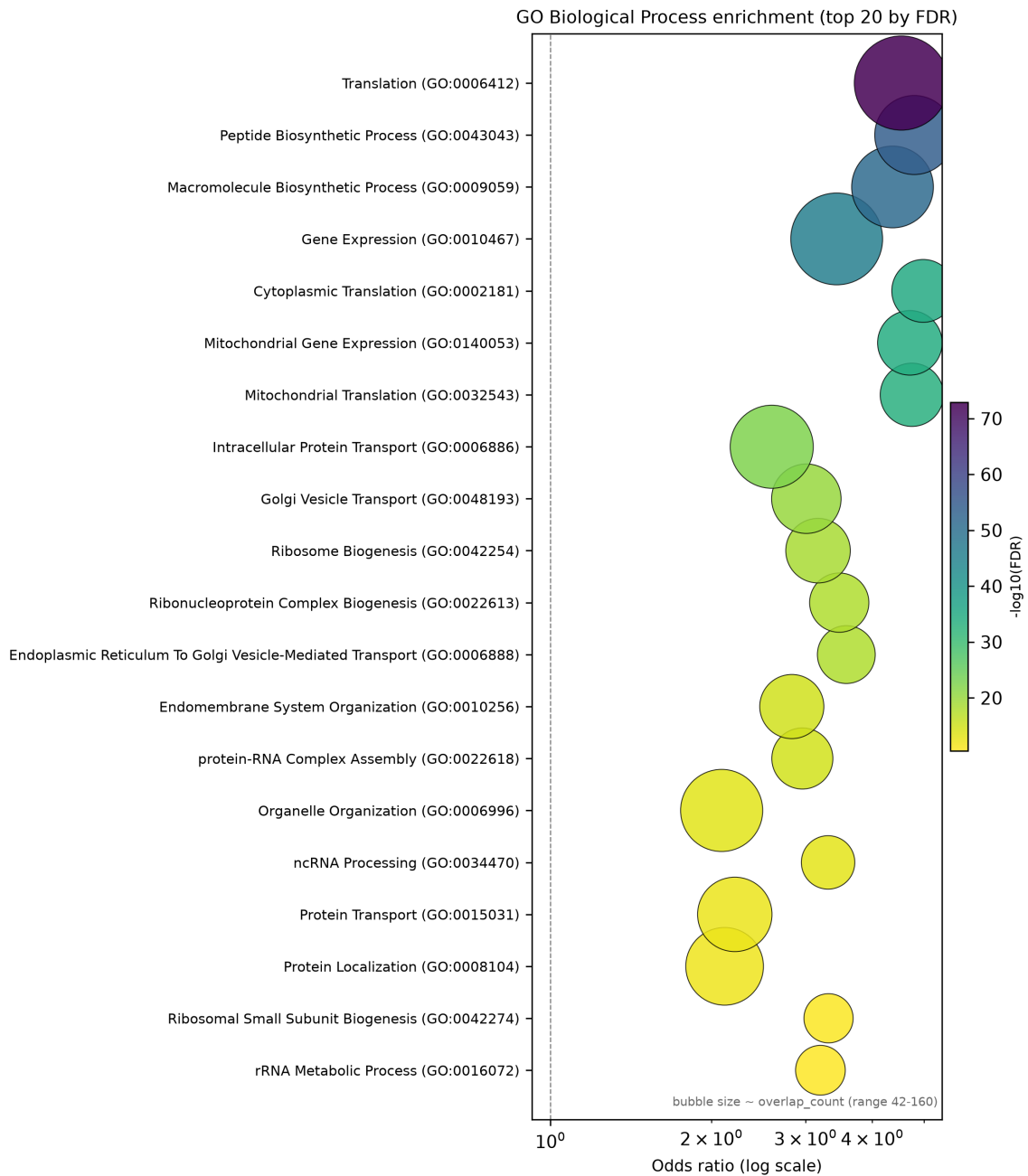


Figure 5: **Gene Ontology Biological Process enrichment of pathogen-target genes.** Top-20 enriched BP terms shown as a bubble plot; bubble size is proportional to the overlap gene count and colour encodes the Benjamini–Hochberg FDR. The dominant signal is host translation and macromolecule biosynthesis.

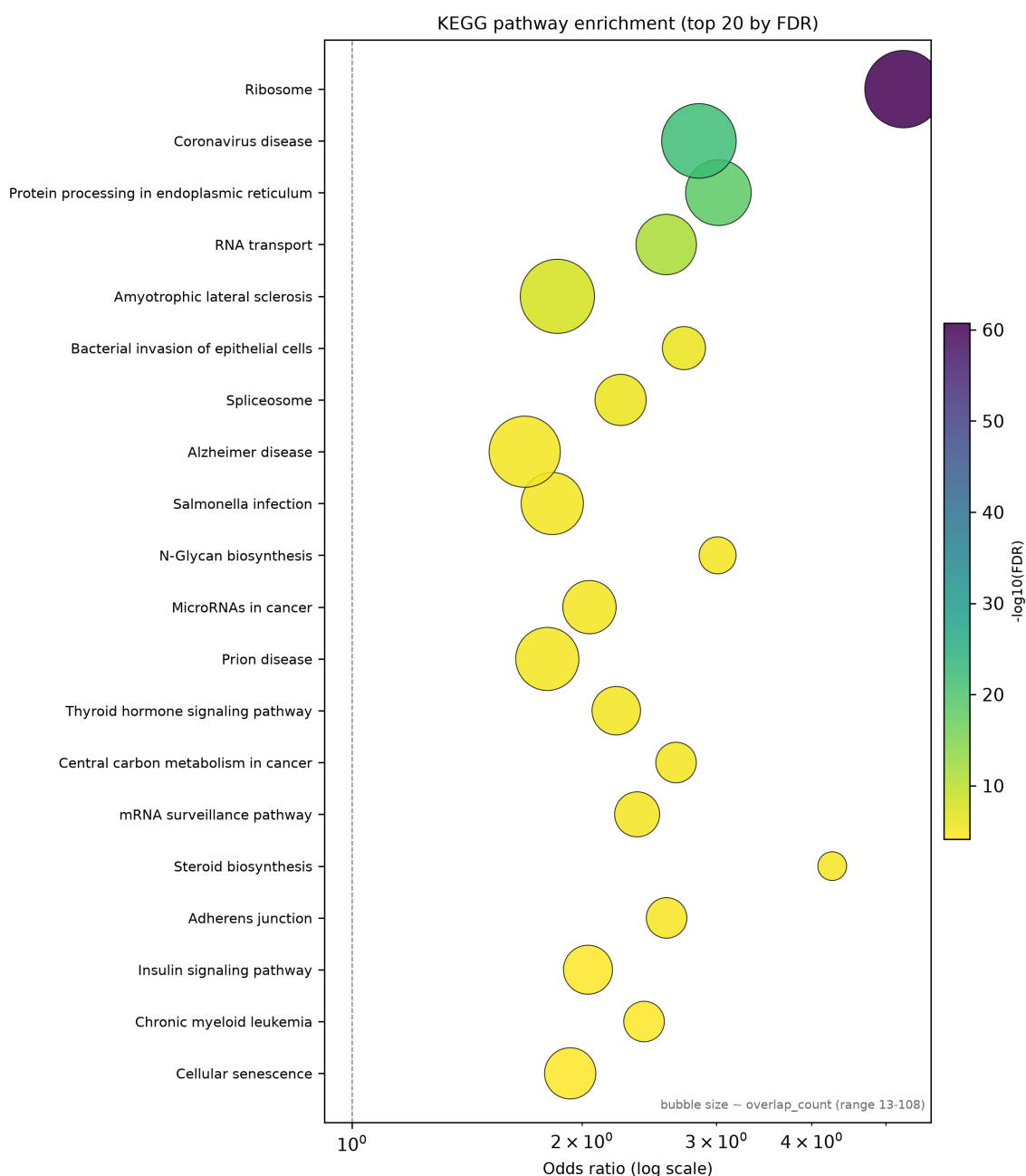


Figure 6: **KEGG pathway enrichment of pathogen-target genes.** Top-20 enriched KEGG pathways. *Ribosome* is the single most enriched pathway (OR = 5.28), followed by coronavirus disease, protein processing in the ER, RNA transport, and infection-associated pathways.

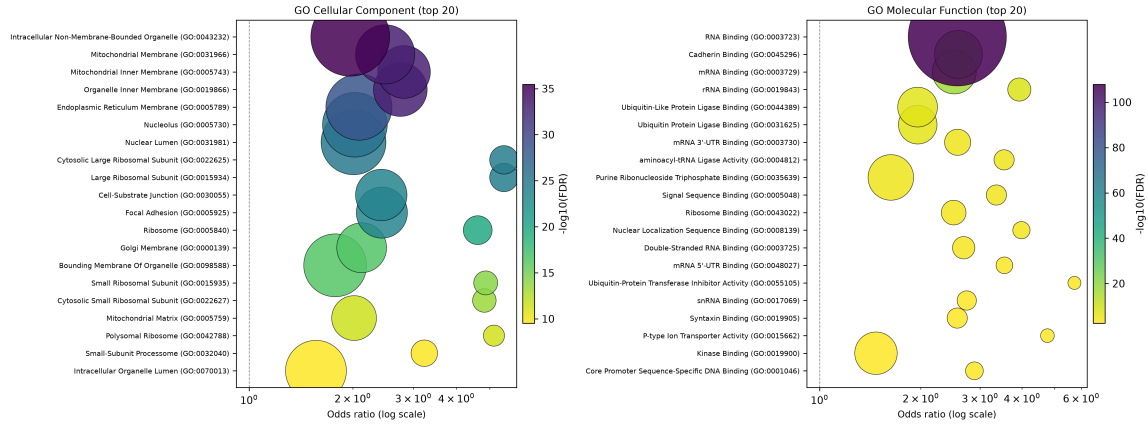


Figure 7: **Gene Ontology Cellular Component and Molecular Function enrichment of pathogen-target genes.** Top-20 terms for each category. CC terms (top) centre on intracellular non-membrane-bounded organelles, mitochondrial/ER membranes, and the nucleolus; MF terms (bottom) are dominated by RNA binding and related nucleic-acid-binding functions.

2.4 A shared core of host proteins is targeted by multiple pathogens

Defining the “core” host target set as proteins targeted by ≥ 2 distinct pathogens yielded 933 proteins, or 35.2% of all 2,649 host targets. Tightening the threshold progressively concentrated the set: 244 proteins (9.2%) were targeted by ≥ 3 pathogens and 30 proteins (1.1%) by ≥ 5 pathogens (Figure 8). Of the 933 core proteins, 866 (92.8%) were present in the STRING backbone, matching the overall PHISTO coverage. The most broadly targeted host protein was DLG1, engaged by 25 distinct pathogens—spanning human papillomaviruses, adenoviruses, herpesviruses, influenza A virus, SARS coronaviruses, and others—followed by the stress-granule components G3BP1 (16 pathogens) and SCRIB (16 pathogens), G3BP2 (13), and the cell-polarity and proteostasis factors ERBIN (11), EZR (11), PPIE (11), PPP1CA (10), and SETD2 (10). This hierarchy of multi-pathogen targeting implicates a convergent set of host cell-organisation, mRNA-stress-granule, and signalling-scaffold proteins as universal contact points for pathogen exploitation.

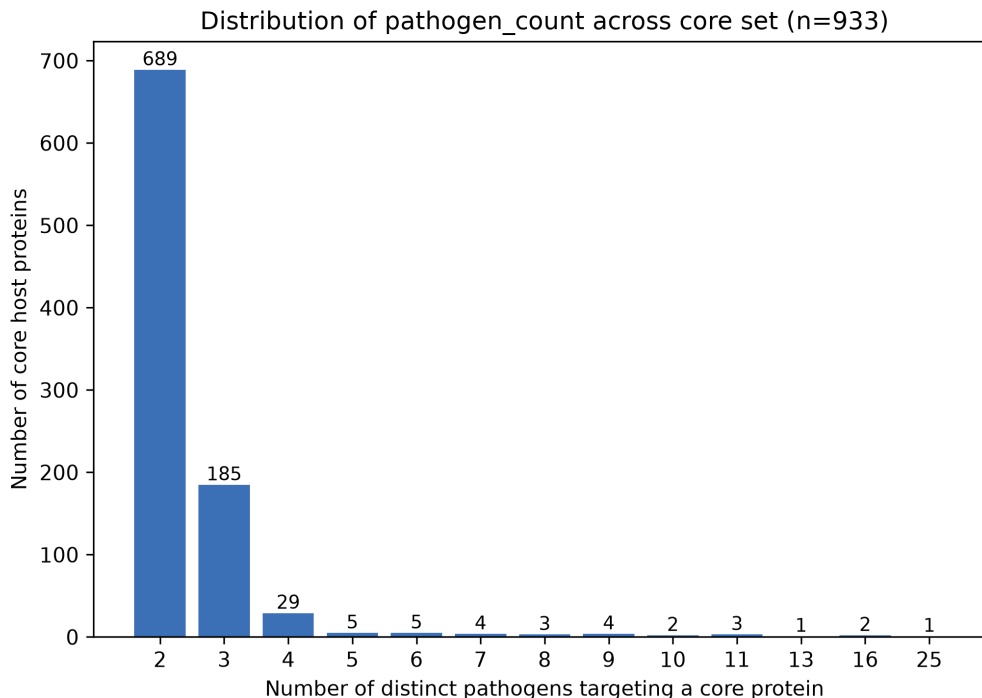


Figure 8: **Distribution of pathogen count across the core host target set.** Most core proteins are targeted by 2–3 pathogens, with a long tail of broadly targeted proteins; DLG1 (25 pathogens) is the most universally targeted host protein. The core set (≥ 2 pathogens) comprises 933 proteins; ≥ 3 yields 244 and ≥ 5 yields 30.

2.5 The top-20 hub targets are dominated by signalling and architectural master regulators

Restricting to pathogen-targeted proteins and ranking by betweenness centrality, the top-20 host hub targets were dominated by master regulators of cell signalling, gene expression, and cell architecture (Table 2, Figure 9, Figure 10). The single most central pathogen target was TP53, with the highest betweenness in the entire network ($C_B = 0.0428$) and the highest degree ($k = 766$), targeted by four pathogens (Hepatitis C virus genotype 1b, SARS coronavirus, hepatitis C virus subtype 1b, and rotavirus sp.) and consequently a core-set member. Other notable hubs included CTNNB1 (β -catenin; $C_B = 0.0284$, $k = 465$), EGFR ($C_B = 0.0207$, $k = 505$; targeted by influenza A virus and *Mycoplasma hyorhinitis*), the actin cytoskeleton component ACTB ($C_B = 0.0199$), the ribosomal component FAU, the amyloid-precursor APP, the inflammatory cytokine TNF, the transcriptional co-activator EP300, the chaperone HSP90AB1, the focal-adhesion protein FN1, the heat-shock protein HSPA8, the chromatin regulator TRIM28, HDAC1, PRKACA, CDH1, the purine-biosynthesis enzyme GART, the AAA-ATPase VCP, TBP, the histone H3C13, and the adhesion receptor CD44.

Six of the top-20 hub targets (TP53, EGFR, ACTB, HSPA8, VCP, and CD44) were core-set members (≥ 2 pathogens). The degree-versus-betweenness scatter (Figure 10) shows that pathogen count and centrality are coupled but not redundant: the most broadly targeted proteins do not always have the highest raw degree, underscoring that betweenness—a measure of bridge-like control over network communication—captures a distinct axis of pathogen-relevant importance. The most enriched GO/KEGG terms attached to each hub (Table 2, last column) recapitulate the global enrichment themes at the individual-protein level: nuclear/nucleolar organisation for TP53, focal-adhesion and cadherin binding for EGFR/ACTB/HSPA8/CD44, RNA binding and translation for FAU/VCP, and ER processing for VCP.

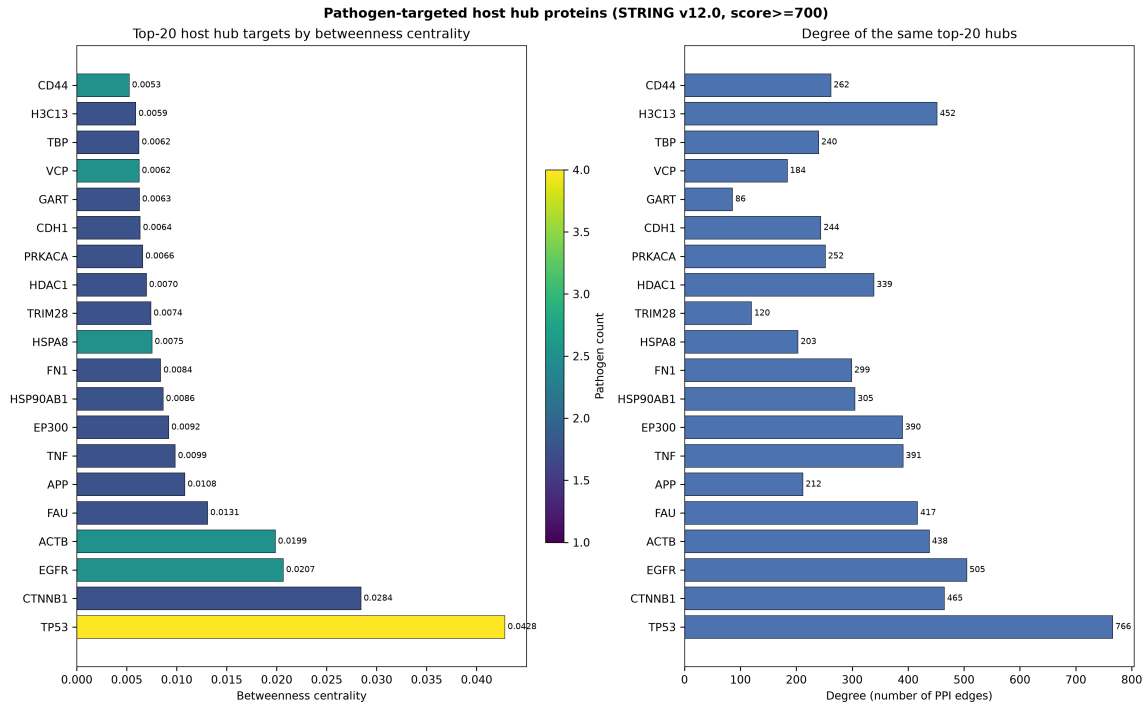


Figure 9: **Top-20 pathogen-targeted host hub proteins ranked by betweenness centrality.** (Top) Betweenness centrality bars, coloured by pathogen count. (Bottom) Degree bars in the same y -order. TP53 is the single most central and highest-degree pathogen target; core-set members (≥ 2 pathogens) are highlighted by warmer colours.

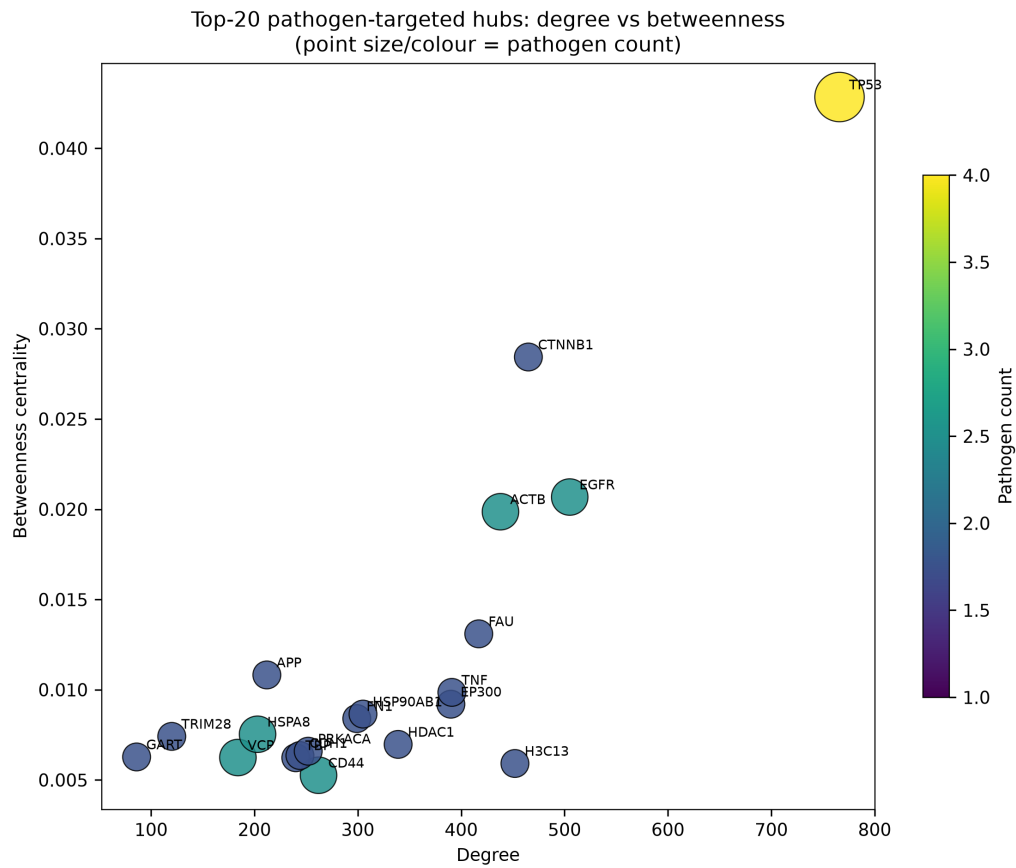


Figure 10: **Degree versus betweenness for the top-20 pathogen-targeted host hubs.** Point size and colour encode the number of distinct pathogens targeting each protein. The two centralities are correlated but not redundant; the most broadly targeted proteins (largest points) cluster toward the upper-right, indicating both high connectivity and high bridge-like control.

Table 2: Top-20 pathogen-targeted host hub proteins ranked by betweenness centrality. *Betw.* = betweenness centrality; *Deg.* = degree; *PC* = number of distinct pathogens targeting the protein; *Core* = member of the core set (≥ 2 pathogens). Top enriched GO/KEGG terms are the three most significant terms *from the global pathogen-target enrichment* that include each protein (per-protein over-representation on a single gene is degenerate).

#	Gene	Betw.	Deg.	PC	Core	Top enriched terms
1	TP53	0.0428	766	4	Yes	Nuclear organelle; Nucleolus; Nuclear lumen
2	CTNNB1	0.0284	465	1	No	Cell-substrate junction; Focal adhesion; Cadherin binding
3	EGFR	0.0207	505	2	Yes	Cell-substrate junction; Focal adhesion; Cadherin binding
4	ACTB	0.0199	438	2	Yes	Nuclear organelle; Cell-substrate junction; Focal adhesion
5	FAU	0.0131	417	1	No	RNA binding; Translation; Ribosome
6	APP	0.0108	212	1	No	Organelle lumen; Regulation of translation
7	TNF	0.0099	391	1	No	Coronavirus disease; ALS; Protein metabolic regulation
8	EP300	0.0092	390	1	No	Protein stabilisation; Nuclear localisation; miRNAs in cancer
9	HSP90AB1	0.0086	305	1	No	RNA binding; Cadherin binding; ER protein processing
10	FN1	0.0084	299	1	No	Organelle lumen; Bacterial invasion; Complex assembly
11	HSPA8	0.0075	203	2	Yes	RNA binding; Cell-substrate junction; Focal adhesion
12	TRIM28	0.0074	120	1	No	RNA binding; Ubiquitin-ligase binding
13	HDAC1	0.0070	339	1	No	miRNAs in cancer; Thyroid signalling; Apoptotic regulation
14	PRKACA	0.0066	252	1	No	RNA processing; mRNA processing; Ubiquitin-ligase binding
15	CDH1	0.0064	244	1	No	Cadherin binding; trans-Golgi; Bacterial invasion
16	GART	0.0063	86	1	No	AMP/GMP/IMP biosynthesis
17	VCP	0.0062	184	2	Yes	RNA binding; Nuclear organelle; ER membrane
18	TBP	0.0062	240	1	No	Spinocerebellar ataxia; HTLV-1 infection; Promoter DNA binding
19	H3C13	0.0059	452	1	No	Shigellosis; NET formation; Alcoholism
20	CD44	0.0053	262	2	Yes	Cell-substrate junction; Focal adhesion; Organelle membrane

2.6 Coronaviruses show the strongest overlap with the shared core target set

Across the 156 pathogens scored, the Jaccard similarity between each pathogen’s target set and the 933-protein core set spanned from 0 to 0.402 (mean 0.013, median 0.002; 150 pathogens had $J > 0$ and 6 had $J = 0$; Figure 11). The pathogen with by far the greatest overlap was SARS coronavirus ($J = 0.4018$; 407 of its 487 targets in the core set), followed closely by SARS-CoV-2 ($J = 0.3927$; 507 of 865), Influenza A virus ($J = 0.2318$; 378 of 1,076), and Zika virus ($J = 0.2037$; 223 of 385). The remaining high-overlap pathogens were human coronaviruses NL63 and 229E, Hepatitis C virus genotype 2a, *Mycobacterium tuberculosis*, and human T-lymphotropic viruses 1 and 2. The clustering of SARS coronaviruses and influenza A at the top reflects both the comparatively large, well-characterised interactomes available for these pathogens and a genuinely convergent strategy of engaging the shared host machinery identified by the core set [9, 25]. At the opposite extreme, pathogens with $J \approx 0$ —such as *Mycoplasma* species, *Haemophilus influenzae*, and hepatitis B virus subtype adw2—target host proteins almost entirely outside the core, indicative of more idiosyncratic, pathogen-specific host-interaction programmes and representing candidate pathogen-specific intervention points.

Table 3: Per-pathogen Jaccard overlap with the 933-protein core host target set. The top-12 pathogens by Jaccard (strongest overlap) and selected low-overlap pathogens are shown; $|T_p|$ = pathogen target count; $|T_p \cap C|$ = intersection with the core set. Across all 156 pathogens scored, the mean Jaccard was 0.013 (median 0.002); 150 pathogens had $J > 0$ and 6 had $J = 0$.

Rank	Pathogen	$ T_p $	$ T_p \cap C $	Jaccard
1	SARS coronavirus	487	407	0.4018
2	SARS-CoV-2	865	507	0.3927
3	Influenza A virus	1,076	378	0.2318
4	Zika virus	385	223	0.2037
5	Human coronavirus NL63	115	102	0.1078
6	Human coronavirus 229E	123	82	0.0842
7	Hepatitis C virus genotype 2a	140	74	0.0741
8	<i>Mycobacterium tuberculosis</i>	184	65	0.0618
9	Human T-lymphotropic virus 1	83	42	0.0431
10	Human T-lymphotropic virus 2	60	33	0.0344
11	Alkhumra hemorrhagic fever virus	38	26	0.0275
12	Tick-borne encephalitis virus	33	24	0.0255
<i>Low-overlap pathogens ($J \approx 0$):</i>				
–	Helicobacter pylori NCTC 11637	7	1	0.0011
–	Shigella flexneri	2	1	0.0011
–	Mycoplasma pulmonis	2	0	0.0000
–	Haemophilus influenzae	2	0	0.0000
–	Hepatitis B virus subtype adw2	1	0	0.0000

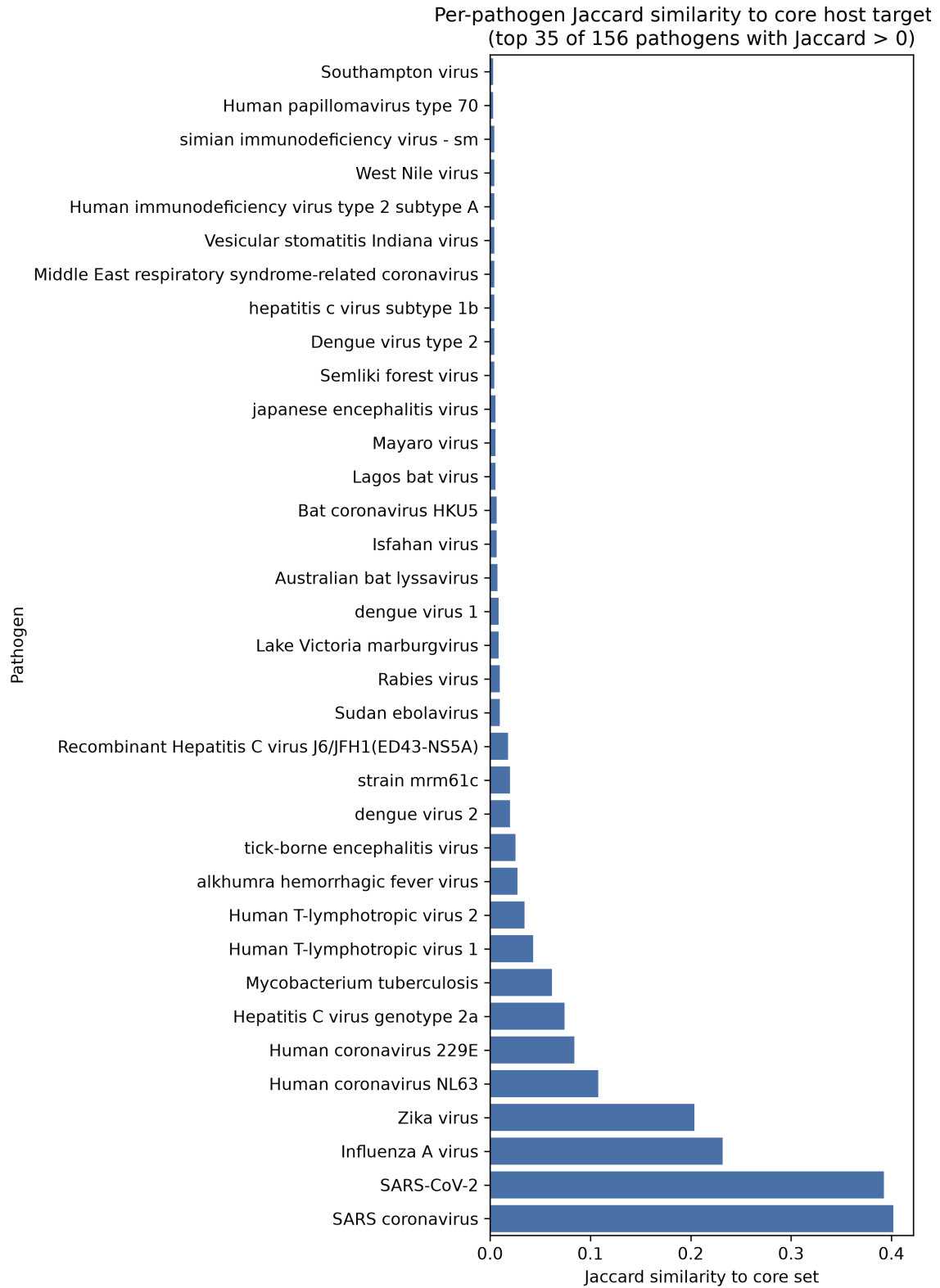


Figure 11: **Per-pathogen Jaccard overlap with the core host target set.** Top-35 pathogens by Jaccard similarity ($J > 0$). SARS coronavirus ($J = 0.402$) and SARS-CoV-2 ($J = 0.393$) show the strongest overlap, indicating their host target repertoires are dominated by shared, multi-pathogen core targets; low-overlap pathogens engage host proteins largely outside the core.

3 Discussion

3.1 Pathogens converge on the most central hubs of the human interactome

The central finding of this study is that pathogen-targeted host proteins are not a random draw from the human proteome but are significantly and reproducibly enriched among the highest-degree and highest-betweenness nodes of the STRING v12.0 PPI backbone. Pathogen targets had a median degree of 28 versus 12 for non-targets (Mann–Whitney $p = 1.74 \times 10^{-146}$, Cliff’s $\delta = 0.326$) and a comparably elevated betweenness ($p = 5.92 \times 10^{-126}$). Crucially, the enrichment persisted under a degree-matched permutation null model that preserved the targets’ degree distribution: even when compared only to non-targets of the same degree, pathogen targets were 1.35-fold more likely to fall in the top-5% degree hubs and 1.19-fold more likely to fall in the top-5% betweenness hubs (empirical $p = 0.0010$ for both). This degree-controlled result is important because it rules out the simplest alternative explanation—that the enrichment is merely an artefact of targets being, on average, more connected—and instead shows that pathogens preferentially engage the very apex of the centrality hierarchy within comparable degree strata. The pattern generalises, across 156 viruses, bacteria, and protozoa, the earlier virus-centric observations of preferential hub targeting [6, 12] and is consistent with the broader centrality–lethality principle [14, 15]: the same topological property that makes a host protein essential for the cell also makes it an efficient lever for a pathogen seeking maximum network reach with minimum interaction effort.

3.2 TP53 and the signalling–architectural core of pathogen targeting

At the apex of the ranked hub list sits TP53, the most connected and most central protein in the entire human PPI backbone (degree 766, betweenness 0.0428) and a target of four pathogens including SARS coronavirus and hepatitis C virus. TP53 is the canonical “guardian of the genome,” a transcription-factor hub governing cell-cycle arrest, apoptosis, and DNA-damage responses, and its inactivation by viral oncoproteins—notably the HPV E6 protein [26–28] and the HCV core protein—is one of the best-characterised mechanisms of pathogen-mediated oncogenesis, and p53 more broadly functions as a host defence factor against diverse viral infections [29]. Its prominence as the top pathogen-targeted hub is therefore both biologically expected and quantitatively striking: pathogens converge on TP53 not because it is merely highly expressed but because, as the most central node in the network, controlling it grants outsized influence over cell-fate decisions. The wider top-20 list reinforces a coherent picture of pathogen-relevant centrality: signalling master regulators (EGFR, CTNNB1, EP300, PRKACA, TNF, HDAC1, TRIM28), cytoskeletal and adhesion scaffolds (ACTB, CDH1, FN1, CD44, EZR), the chaperone machinery (HSP90AB1, HSPA8, VCP), the translational apparatus (FAU, H3C13), and the amyloid-precursor APP. The recurrence of focal-adhesion and cadherin-binding terms among the top enriched functions of these individual hubs aligns with the established importance of membrane, cytoskeleton, and cell–matrix remodelling in pathogen entry, spread, and immune evasion [6, 13, 30, 31].

3.3 Functional convergence on translation, proteostasis, and the secretory pathway

The GO/KEGG enrichment provides a functional readout of the same hub-targeting logic. The single dominant theme—translation and the ribosomal machinery (top GO BP term: Translation; top KEGG pathway: Ribosome, OR 5.28)—reflects the obligate dependence of viruses and many intracellular bacteria on host protein synthesis [23, 24, 32]. Pathogens must commandeer host ribosomes, translation factors, and tRNA charging to produce their own proteins, and

the ribosomal proteins and translation factors are accordingly among the most heavily represented pathogen targets. A second, tightly linked cluster concerns proteostasis and the secretory pathway: protein processing in the endoplasmic reticulum, intracellular protein transport, Golgi vesicle transport, *N*-glycan biosynthesis, the spliceosome, and RNA transport. These reflect pathogen reliance on the host ER folding, glycosylation, and vesicular-trafficking infrastructure to mature and traffic pathogen or effector proteins, and to remodel secretory pathways for replication [33, 34]. The enrichment of infection-relevant KEGG pathways—coronavirus disease, Salmonella infection, bacterial invasion of epithelial cells—and neurodegeneration pathways sharing NF- κ B, PI3K-AKT, cytoskeletal, and proteasome components confirms that the target set is concentrated in the host modules that pathogens actually engage, and that the network and functional signals are mutually consistent [35–38].

3.4 A shared core and the universal contact points

Defining the core as host proteins targeted by ≥ 2 pathogens yielded 933 proteins, a set that narrows sharply to 244 (≥ 3) and 30 (≥ 5) as the threshold tightens. The most broadly targeted core protein, DLG1, was engaged by 25 distinct pathogens—a striking convergence given that these span papillomaviruses, adenoviruses, herpesviruses, influenza A, and SARS coronaviruses. DLG1 is a PDZ-domain scaffolding protein of the MAGUK family that organises cell polarity, adherens junctions, and signalling complexes, and is a documented target of viral oncoproteins (e.g. HPV E6, adenovirus E4-ORF1) that disrupt epithelial polarity and tight-junction integrity [39, 40]. The co-dominance of stress-granule components (G3BP1, G3BP2, SCRIB) and cell-polarity factors (ERBIN, EZR) in the core further implicates mRNA homeostasis and cell-architecture remodelling as universal, cross-pathogen contact points [41, 42]. The fact that the same handful of host proteins is reached by phylogenetically unrelated pathogens is strong evidence for convergent evolution toward a shared set of network-vulnerable hubs, and it identifies these proteins as priority candidates for understanding—and ultimately therapeutically intervening in—universal pathogen strategies.

3.5 Coronaviruses dominate the core-overlap landscape

The per-pathogen Jaccard analysis revealed a wide spectrum of overlap with the core set, from SARS coronavirus ($J = 0.402$) and SARS-CoV-2 ($J = 0.393$) down to a tail of pathogens with $J \approx 0$. Two factors drive the coronaviral dominance. First, SARS-CoV and SARS-CoV-2 are among the pathogens with the largest, most systematically mapped host interactomes in the dataset [9, 25, 43, 44], so their high Jaccard partly reflects interactome depth. Second, and more substantively, a high Jaccard indicates that a pathogen’s targets are drawn predominantly from the shared core, i.e. that it relies on the same universal host machinery rather than on idiosyncratic, pathogen-specific contacts. The clustering of influenza A, Zika, and human coronaviruses alongside SARS at the top of the Jaccard ranking supports a model in which well-characterised respiratory and emerging viruses converge on a common host-network module [8, 45–47]. By contrast, the low-Jaccard pathogens—several *Mycoplasma* species, *Haemophilus influenzae*, and hepatitis B virus—target host proteins largely outside the core, reflecting more specialised or niche-specific interaction strategies; these proteins, being less shared, are candidates for pathogen-specific intervention points with a lower risk of host-network-wide perturbation.

3.6 Limitations

Several limitations should temper the interpretation of these results. First, the analysis is only as complete as the underlying data: PHISTO, though curated, is biased toward pathogens whose host interactions have been intensively studied—predominantly viruses of biomedical

importance—so the high Jaccard of coronaviruses and influenza partly reflects research effort. Second, the STRING confidence cutoff (≥ 0.7) trades coverage for reliability; a lower threshold would admit more—but noisier—edges and could shift centrality estimates, particularly for mid-ranked nodes. Third, betweenness centrality was estimated by approximate sampling ($k = 1,000$ source nodes, seed 42) rather than the exact Brandes algorithm; while top-hub membership (e.g. TP53, CTNNB1) is stable, mid-range betweenness rankings are seed-dependent and should be interpreted as estimates. Fourth, the core-set definition (≥ 2 pathogens) is necessarily a compromise: the size falls from 933 to 30 across the ≥ 2 , ≥ 3 , ≥ 5 thresholds, and downstream Jaccard values scale with this choice; the sensitivity analysis at ≥ 3 and ≥ 5 provides bounds but the “true” core boundary is inherently fuzzy. Fifth, the network analysed is static and unweighted; pathogen–host interactions occur dynamically in time, tissue, and subcellular compartment, and a single aggregated graph cannot capture this context. Sixth, ORA on a large foreground set (15% of the STRING universe) is highly powered, so many broad GO terms reach significance trivially; the specificity and odds ratios of the top terms (ribosome, translation, ER processing) are the meaningful signal, not the raw count of significant terms. Seventh, and most importantly, the hub-targeting enrichment is subject to a shared study-bias (detection) confound: proteins studied intensively accrue both more STRING edges and more recorded pathogen–host interactions, so high degree and high target-recording probability have a common cause. The degree-matched null controls for degree but not for the underlying study bias that degree proxies; within a degree decile, pathogen targets may still be the more-studied subset, so the residual $1.35\times$ enrichment could in principle reflect residual study bias rather than biological hub-preference. A definitive separation would require restricting to proteins matched on a study-count proxy or a study-bias-corrected binary-only interactome, which is beyond the present design. Eighth, the per-pathogen Jaccard-overlap-with-core analysis is partially circular: the core set is defined from the same pathogen–target data, so pathogens with large target counts that helped populate the core (e.g. SARS-CoV-2, 865 targets) mechanically score higher Jaccard independent of biological convergence. A leave-one-out Jaccard, in which the core is recomputed excluding pathogen p before scoring $J(p)$, would remove this self-contribution; the present ranking should therefore be interpreted as descriptive of overlap rather than as evidence of independent convergence.

3.7 Future directions

These limitations point to clear next steps. Dynamic and condition-specific networks—integrating tissue, temporal, and subcellular resolution—would refine the hub-targeting picture beyond the static aggregate. Weighted and directed analyses, including network-flow and controllability measures, could quantify not only *whether* pathogens target hubs but *how much* network control those contacts confer. Experimental validation of the predicted core contact points—particularly DLG1, G3BP1/2, and the top-20 hubs—through targeted perturbation would test the causal predictions of the centrality–targeting model. Finally, the shared core identified here offers a translational lever: host-directed therapies aimed at the most universally targeted hubs carry the broadest spectrum but also the greatest toxicity risk, whereas pathogen-specific, low-Jaccard targets offer narrower, potentially safer intervention points. Distinguishing these two classes systematically is a natural extension of the per-pathogen Jaccard framework.

4 Methods

This study followed a six-step computational pipeline (Figure 1) that integrates a curated human protein–protein interaction (PPI) backbone with a large, literature-curated catalogue of pathogen–host protein interactions, annotates every host protein with network topological descriptors, statistically tests whether pathogen targets are enriched among network hubs, function-

ally characterises the target set, and quantifies the overlap of each pathogen’s target repertoire with a shared “core” host target set. All analyses were performed in Python using the NetworkX library [48]. A single random seed (`random_state=42`) was used for every stochastic procedure to ensure full reproducibility.

4.1 Data acquisition

Human PPI backbone. The human protein–protein interaction network was retrieved from STRING v12.0 [20, 49–51], the current production version of a continuously updated resource whose confidence-scoring, evidence-integration, and programmatic-access framework has developed across successive releases [52, 53]. The full `9606.protein.links` table was downloaded from the STRING FTP repository and filtered to retain only interactions with a combined confidence score of at least 700 (equivalent to confidence ≥ 0.7), encompassing all evidence types (neighbourhood, fusion, co-occurrence, co-expression, experiments, databases, text-mining). STRING protein identifiers were mapped to HGNC gene symbols using the STRING alias file (`9606.protein.aliaes`), producing a canonical, gene-symbol-labelled edge list. A quality re-audit of the deduplicated edge file identified and removed 11 residual self-loops and 441 duplicate canonical (undirected) pairs, yielding a true unique undirected edge set that was loaded as an undirected graph.

Pathogen–host interactions. Pathogen–host protein interactions were obtained from the curated catalogue accompanying Ahmed et al. [5], a manually assembled resource of experimentally validated pathogen–host interactions aggregated from IntAct, BioGRID, and the primary literature. The Supplementary Tables (MOESM3–5) accompanying that publication—archived at the PHISTO repository (<http://www.phisto.org>)—were used as the source of host–pathogen edges. Records were filtered to human host proteins, deduplicated, and the host protein identifiers (predominantly UniProt accessions) were mapped to HGNC gene symbols via the UniProt-to-gene cross-reference table, matching them against the STRING backbone. The final cleaned pathogen–host interaction table contained 5,502 interaction records spanning 156 distinct pathogens (viral, bacterial, and protozoan) and 2,649 unique host proteins.

4.2 Network construction and topological annotation

The STRING backbone was loaded as an undirected graph $G = (V, E)$ using NetworkX. For every node $v \in V$ the following descriptors were computed: (i) *degree* $k(v)$ and *degree centrality* $C_D(v) = k(v)/(|V| - 1)$; (ii) *betweenness centrality*,

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where σ_{st} is the number of shortest paths between s and t and $\sigma_{st}(v)$ the number passing through v . Because exact Brandes betweenness [19] on a 16,059-node graph is computationally expensive, betweenness was estimated by the NetworkX approximate-sampling variant with $k = 1,000$ source nodes (`random_state=42`) and normalised by the undirected-graph factor $2/[(n-1)(n-2)]$. Each node was annotated with a binary flag indicating whether it is a pathogen target (i.e., present in the PHISTO host set and connected to at least one pathogen) and, for targets, the number and identity of distinct pathogens targeting it. The node annotation table thus records, for every host protein, its degree, degree centrality, betweenness centrality, target status, and pathogen count.

4.3 Topological enrichment testing

We tested the hypothesis that pathogen targets are enriched among high-degree and high-betweenness hubs using two complementary, non-parametric approaches.

Mann–Whitney U test. The degree and betweenness distributions of pathogen targets ($n = 2,453$) and non-target proteins ($n = 13,606$) were compared with a one-sided Mann–Whitney U test [54] (alternative: targets have greater centrality than non-targets), with two-sided p -values also reported. The effect size was quantified as Cliff’s δ [55], computed directly from the U statistic as

$$\delta = \frac{2U - n_1n_2}{n_1n_2},$$

where n_1 and n_2 are the target and non-target group sizes. Magnitude was interpreted following the thresholds $|\delta| < 0.147$ negligible, 0.147 – 0.33 small, 0.33 – 0.474 medium, > 0.474 large.

Degree-matched permutation null model. Because the Mann–Whitney test compares entire distributions, a significant result could in principle arise solely because pathogen targets have higher degree on average, rather than because they cluster at the very top of the centrality hierarchy. To disentangle these, we constructed a degree-matched null model [56]. Nodes were partitioned into ten degree-decile bins. Within each bin, the number of pathogen targets was counted, and an equal number of non-target nodes was sampled uniformly from the same bin (`random_state=42`); pooling across bins yields a random set whose degree distribution matches the target set exactly but whose hub-membership is random under the null. The test statistic was the proportion of the set falling in the top-5% hubs (computed separately for degree, threshold = 111, and for betweenness, threshold = 0.000743). The procedure was repeated 1,000 times; the empirical p -value was $p = (\#\{\text{perm} \geq \text{observed}\} + 1) / (n_{\text{perm}} + 1)$.

4.4 Functional enrichment analysis

Gene Ontology (GO) and KEGG pathway enrichment of the pathogen-target gene set was performed with a custom hypergeometric over-representation analysis (ORA) [57, 58]. The foreground set was the 2,453 pathogen-target genes present in the STRING graph; the background (universe) was the 16,059 genes of the STRING backbone, so that the foreground is a strict subset of the background (verified). Gene-set libraries were obtained from Enrichr [59]: GO Biological Process 2023 (5,406 terms), GO Cellular Component 2023 (472 terms), GO Molecular Function 2023 (1,147 terms), and KEGG 2021 Human (320 terms). GO term definitions and the GO framework follow the Gene Ontology Consortium [60–62], and KEGG pathways follow Kanehisa and Goto [63, 64]. Because Enrichr’s online API applies a fixed per-library background that does not equal the STRING universe, we fetched only the gene sets and computed the one-tailed hypergeometric test ourselves against the 16,059-gene STRING background:

$$P(X \geq k) = \text{hypergeom.sf}(k - 1; M, K, n),$$

where M is the background size, K the term size within the background, n the foreground size, and k the foreground–term overlap. The odds ratio was computed as $(k/n)/(K/M)$. Benjamini–Hochberg false-discovery-rate (FDR) correction [65] was applied within each category (BP, CC, MF, KEGG), with significance at $\text{FDR} < 0.05$.

4.5 Core host target set and per-pathogen overlap

For each host protein we counted the number of *distinct* pathogens targeting it (`nunique`, so repeated records of the same pathogen–host pair count once). The shared “core” host target

set C was defined as host proteins targeted by ≥ 2 distinct pathogens; as a sensitivity analysis, core-set sizes were also recorded at thresholds ≥ 3 and ≥ 5 pathogens. For each pathogen p with target set T_p , the Jaccard similarity to the core set was

$$J(p) = \frac{|T_p \cap C|}{|T_p \cup C|}.$$

Pathogens with empty intersection ($J = 0$) were retained in the results table. A core protein was annotated “in STRING” if its gene symbol was present in the STRING backbone node annotation table.

4.6 Reproducibility

All stochastic operations used `random_state=42`; betweenness was estimated with $k = 1,000$ sampled source nodes (seed 42); the permutation null model used 1,000 permutations (seed 42). Every parameter, threshold, and intermediate file is documented in the analysis manifest. The STRING confidence cutoff (700), the core-set threshold (≥ 2 pathogens), and the FDR threshold (0.05) were pre-specified before testing.

Data and code availability

The STRING v12.0 human PPI backbone (`9606.protein.links`, `9606.protein.aliaes`) is publicly available from the STRING repository (<https://string-db.org>). The pathogen–host interaction data (Supplementary Tables MOESM3–5 of Ahmed et al. [5]) are publicly available and archived at the PHISTO repository (<http://www.phisto.org>). GO 2023 and KEGG 2021 gene-set libraries were obtained from Enrichr [59] on the run date. All analysis scripts (Python/NetworkX), the cleaned interaction table, the node annotation table, and the full results files are provided in the project repository accompanying this manuscript; every stochastic step is seeded (`random_state=42`) for exact reproducibility.

5 Conclusion

By integrating the STRING v12.0 human protein–protein interaction backbone with a curated catalogue of $\sim 5,500$ pathogen–host interactions spanning 156 pathogens, we show that pathogens—viral, bacterial, and protozoan—do not engage the host proteome at random but preferentially target the most topologically central hubs of the human interactome. This enrichment is significant by distributional comparison (Mann–Whitney $p \leq 10^{-126}$) and, critically, survives a degree-matched permutation null model, demonstrating that pathogens reach for the apex of the centrality hierarchy within comparable degree strata rather than merely benefiting from higher average connectivity. The functional consequence of this hub-targeting logic is a striking convergence on a shared module of host cell biology: translation and the ribosome, proteostasis, and the secretory pathway. A core set of 933 host proteins targeted by multiple pathogens—narrowing to 30 proteins targeted by five or more—identifies the universal contact points of pathogen exploitation, with DLG1 (25 pathogens) at its apex and TP53 as the single most central and most connected pathogen target. SARS coronavirus draws its host repertoire most heavily from this core ($J = 0.402$), but the broader picture is one of cross-pathogen convergence on a common network-vulnerable hub set. These findings refine the centrality–lethality principle into a centrality–targeting principle for infectious disease and delineate a tiered therapeutic landscape: broad-spectrum, host-directed strategies aimed at the shared core carry the widest applicability but also the greatest toxicity, whereas low-Jaccard, pathogen-specific targets offer narrower and potentially safer intervention points. The network-analytic framework

presented here—degree-controlled enrichment testing, multi-threshold core-set definition, and per-pathogen Jaccard overlap—provides a general, reproducible template for interrogating how the diversity of the microbial world converges on, and diverges from, a common logic of host-network exploitation.

References

- [1] J F Rual, K Venkatesan, T Hao, T Hirozane-Kishikawa, A Dricot, N Li, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437(7062): 1173–1178, 2005. doi: 10.1038/nature04209.
- [2] T Rolland, M Taşan, B Charloteaux, S J Pevzner, Q Zhong, N Sahni, et al. A proteome-scale map of the human interactome network. *Cell*, 159(5):1212–1226, 2014. doi: 10.1016/j.cell.2014.10.050.
- [3] U Stelzl, U Worm, M Lalowski, C Haenig, F H Brembeck, H Goehler, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6): 957–968, 2005. doi: 10.1016/j.cell.2005.08.029.
- [4] K Venkatesan, J F Rual, A Vazquez, U Stelzl, I Lemmens, T Hirozane-Kishikawa, et al. An empirical framework for binary interactome mapping. *Nature Methods*, 6(1):83–90, 2009. doi: 10.1038/nmeth.1280.
- [5] H Ahmed, T C Howton, Y Sun, N Weinberger, Y Belkhadir, and M S Mukhtar. Network biology discovers pathogen contact points in host protein-protein interactomes. *Nature Communications*, 9:2312, 2018. doi: 10.1038/s41467-018-04632-8.
- [6] B de Chassey, V Navratil, L Tafforeau, M S Hiet, A Aublin-Gex, et al. Hepatitis C virus infection protein network. *Molecular Systems Biology*, 4:230, 2008. doi: 10.1038/msb.2008.66.
- [7] N Simonis, J F Rual, I Lemmens, M Boxus, T Hirozane-Kishikawa, et al. Host-pathogen interactome mapping for HTLV-1 and -2 retroviruses. *Retrovirology*, 9:26, 2012. doi: 10.1186/1742-4690-9-26.
- [8] S D Shapira, I Gat-Viks, B O Shum, A Dricot, M M de Grace, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell*, 139(7):1255–1267, 2009. doi: 10.1016/j.cell.2009.12.018.
- [9] D E Gordon, G M Jang, M Bouhaddou, J Xu, K Obernier, K M White, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583(7816):459–468, 2020. doi: 10.1038/s41586-020-2286-9.
- [10] T Ideker and R Sharan. Protein networks in disease. *Genome Research*, 18(4):644–652, 2008. doi: 10.1101/gr.071852.107.
- [11] A-L Barabási, N Gulbahce, and J Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011. doi: 10.1038/nrg2918.
- [12] P Devkota, M C Danzi, and S Wuchty. Beyond degree and betweenness centrality: alternative topological measures to predict viral targets. *PLoS ONE*, 13(5):e0197595, 2018. doi: 10.1371/journal.pone.0197595.
- [13] G Lasso, S V Mayer, E R Winkelmann, T Chu, O Elliot, et al. A structure-informed atlas of human-virus interactions. *Cell*, 178(6):1526–1541, 2019. doi: 10.1016/j.cell.2019.08.005.

- [14] H Jeong, S P Mason, A-L Barabási, and Z N Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001. doi: 10.1038/35075138.
- [15] X He and J Zhang. Why do hubs tend to be essential in protein networks? *PLoS Genetics*, 2(6):e88, 2006. doi: 10.1371/journal.pgen.0020088.
- [16] E Zotenko, J Mestre, D P O’Leary, and T M Przytycka. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Computational Biology*, 4(8):e1000140, 2008. doi: 10.1371/journal.pcbi.1000140.
- [17] K Raman, N Damaraju, and G K Joshi. The organisational structure of protein networks: revisiting the centrality-lethality hypothesis. *Systems and Synthetic Biology*, 8(1):73–81, 2013. doi: 10.1007/s11693-013-9123-5.
- [18] B de Chassey, L Meyniel-Schicklin, J Vonderscher, P André, and V Lotteau. Virus-host interactomics: new insights and opportunities for antiviral drug discovery. *Genome Medicine*, 6(11):115, 2014. doi: 10.1186/s13073-014-0115-1.
- [19] U Brandes. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163–177, 2001. doi: 10.1080/0022250X.2001.9990249.
- [20] D Szklarczyk, R Kirsch, M Koutrouli, K Nastou, F Mehryary, R Hachilif, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, 2023. doi: 10.1093/nar/gkac1000.
- [21] A-L Barabási and Z N Oltvai. Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004. doi: 10.1038/nrg1272.
- [22] M P Joy, A Brock, D E Ingber, and S Huang. High-betweenness proteins in the yeast protein interaction network. *BioMed Research International*, 2005(2):96–103, 2005. doi: 10.1155/jbb.2005.96.
- [23] D Walsh and I Mohr. Viral subversion of the host protein synthesis machinery. *Nature Reviews Microbiology*, 9(12):860–875, 2011. doi: 10.1038/nrmicro2655.
- [24] N Stern-Ginossar, B Weisburd, A Michalski, V T K Le, M Y Hein, et al. Decoding human cytomegalovirus. *Science*, 338(6110):1088–1093, 2012. doi: 10.1126/science.1227919.
- [25] S Pfefferle, J Schöpf, M Kögl, C C Friedel, M A Müller, et al. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathogens*, 7(10):e1002331, 2011. doi: 10.1371/journal.ppat.1002331.
- [26] B A Werness, A J Levine, and P M Howley. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science*, 248(4951):76–79, 1990. doi: 10.1126/science.2157286.
- [27] M Scheffner, B A Werness, J M Huibregtse, A J Levine, and P M Howley. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*, 63(6):1129–1136, 1990. doi: 10.1016/0092-8674(90)90409-8.
- [28] A Storey, M Thomas, A Kalita, C Harwood, D Gardiol, et al. Role of a p53 polymorphism in the development of human papilloma-virus-associated cancer. *Nature*, 393(6682):229–234, 1998. doi: 10.1038/30400.
- [29] R Aloni-Grinstein, M Charni-Natan, H Solomon, and V Rotter. p53 and the viral connection: back into the future. *Cancers*, 10(6):178, 2018. doi: 10.3390/cancers10060178.

- [30] K Radtke, K Döhner, and B Sodeik. Viral interactions with the cytoskeleton: a hitchhiker’s guide to the cell. *Cellular Microbiology*, 8(3):387–400, 2006. doi: 10.1111/j.1462-5822.2005.00679.x.
- [31] T E B Stradal and M Schelhaas. Actin dynamics in host-pathogen interaction. *FEBS Letters*, 592(22):3658–3669, 2018. doi: 10.1002/1873-3468.13173.
- [32] A E Firth and I Brierley. Non-canonical translation in RNA viruses. *Journal of General Virology*, 93(Pt 7):1385–1409, 2012. doi: 10.1099/vir.0.042499-0.
- [33] R J Kaufman. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes & Development*, 13(10):1211–1233, 1999. doi: 10.1101/gad.13.10.1211.
- [34] T S Fung and D X Liu. Coronavirus infection, ER stress, apoptosis and innate immunity. *Frontiers in Microbiology*, 5:296, 2014. doi: 10.3389/fmicb.2014.00296.
- [35] M S Hayden and S Ghosh. Signaling to NF- κ B. *Genes & Development*, 18(18):2195–2224, 2004. doi: 10.1101/gad.1228704.
- [36] N Frias-Staheli, N V Giannakopoulos, M Kikkert, S L Taylor, A Bridgen, et al. Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host & Microbe*, 2(6):404–416, 2007. doi: 10.1016/j.chom.2007.09.014.
- [37] M A Germain, L Chatel-Chaix, B Gagné, É Bonneil, P Thibault, et al. Elucidating novel hepatitis C virus-host interactions using combined mass spectrometry and functional genomics approaches. *Molecular & Cellular Proteomics*, 13(1):184–203, 2014. doi: 10.1074/mcp.M113.030155.
- [38] B de Chassey, L Meyniel-Schicklin, A Aublin-Gex, V Navratil, T Chantier, P André, and V Lotteau. Structure homology and interaction redundancy for discovering virus-host protein interactions. *EMBO Reports*, 14(10):938–944, 2013. doi: 10.1038/embor.2013.130.
- [39] H Chang, H Wu, P Hou, M Aizaz, R Yang, et al. DLG1 promotes the antiviral innate immune response by inhibiting p62-mediated autophagic degradation of IKK ϵ . *Journal of Virology*, 97(12):e0150123, 2023. doi: 10.1128/jvi.01501-23.
- [40] S Nakagawa, T Yano, K Nakagawa, S Takizawa, Y Suzuki, et al. Analysis of the expression and localisation of a LAP protein, human scribble, in the normal and neoplastic epithelium of uterine cervix. *British Journal of Cancer*, 90(1):194–199, 2004. doi: 10.1038/sj.bjc.6601465.
- [41] L C Reineke and R E Lloyd. Diversion of stress granules and P-bodies during viral infection. *Virology*, 436(2):255–267, 2013. doi: 10.1016/j.virol.2012.11.017.
- [42] N Kedersha, M D Panas, C A Achorn, S M Lyons, S Tisdale, et al. G3BP-Caprin1-USP10 complexes mediate stress granule condensation and associate with 40S subunits. *Journal of Cell Biology*, 212(7):845–860, 2016. doi: 10.1083/jcb.201508028.
- [43] J P Davies, K M Almasy, E F McDonald, and L Plate. Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus nonstructural proteins identifies unique and shared host-cell dependencies. *ACS Infectious Diseases*, 6(12):3174–3189, 2020. doi: 10.1021/acsinfecdis.0c00500.
- [44] X Liu, S Huuskonen, T Laitinen, T Redchuk, M Bogacheva, et al. SARS-CoV-2-host proteome interactions for antiviral drug discovery. *Molecular Systems Biology*, 17(11):e10396, 2021. doi: 10.15252/msb.202110396.

- [45] T Watanabe, E Kawakami, J E Shoemaker, T J Lopes, Y Matsuoka, et al. Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host & Microbe*, 16(6):795–805, 2014. doi: 10.1016/j.chom.2014.11.002.
- [46] P Scaturro, A Stukalov, D Hagstöm, M Goroszka, M Dorner, et al. Global interactomics uncovers extensive organellar targeting by Zika virus. *Molecular & Cellular Proteomics*, 17: 2242–2255, 2018. doi: 10.1074/mcp.TIR118.000800.
- [47] A Karlas, N Machuy, Y Shin, K P Pleissner, A Artarini, et al. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature*, 463(7282):818–822, 2010. doi: 10.1038/nature08760.
- [48] A A Hagberg, D A Schult, and P J Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, pages 11–15, 2008. doi: 10.25080/tcww9851.
- [49] D Szklarczyk, A L Gable, K C Nastou, D Lyon, R Kirsch, S Pyysalo, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49(D1):D605–D612, 2021. doi: 10.1093/nar/gkaa1074.
- [50] D Szklarczyk, A L Gable, D Lyon, A Junge, S Wyder, J Huerta-Cepas, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1):D607–D613, 2019. doi: 10.1093/nar/gky1131.
- [51] D Szklarczyk, J H Morris, H Cook, M Kuhn, S Wyder, M Simonovic, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*, 45(D1):D362–D368, 2017. doi: 10.1093/nar/gkw937.
- [52] D Szklarczyk, A Franceschini, S Wyder, K Forslund, D Heller, J Huerta-Cepas, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43(Database issue):D447–D452, 2015. doi: 10.1093/nar/gku1003.
- [53] A Franceschini, D Szklarczyk, S Frankild, M Kuhn, M Simonovic, A Roth, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(Database issue):D808–D815, 2013. doi: 10.1093/nar/gks1094.
- [54] H B Mann and D R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947. doi: 10.1214/aoms/1177730491.
- [55] N Cliff. Dominance statistics: ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3):494–509, 1993. doi: 10.1037/0033-2909.114.3.494.
- [56] S Maslov and K Sneppen. Protein interaction networks beyond artifacts. *FEBS Letters*, 530(1-3):255–256, 2002. doi: 10.1016/S0014-5793(02)03428-2.
- [57] E I Boyle, S Weng, J Gollub, H Jin, D Botstein, et al. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20(18):3710–3715, 2004. doi: 10.1093/bioinformatics/bth456.
- [58] A Subramanian, P Tamayo, V K Mootha, S Mukherjee, B L Ebert, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. doi: 10.1073/pnas.0506580102.

- [59] M V Kuleshov, M R Jones, A D Rouillard, N F Fernandez, Q Duan, Z Wang, et al. Enrichr: a comprehensive gene set enrichment analysis web-based tool. *Nucleic Acids Research*, 44 (W1):W90–W97, 2016. doi: 10.1093/nar/gkw377.
- [60] M Ashburner, C A Ball, J A Blake, D Botstein, H Butler, J M Cherry, et al. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, 2000. doi: 10.1038/75556.
- [61] The Gene Ontology Consortium. The Gene Ontology resource: enriching a GO database. *Nucleic Acids Research*, 49(D1):D325–D334, 2021. doi: 10.1093/nar/gkaa1113.
- [62] S Aleksander, J P Balhoff, S Carbon, J M Cherry, H Drabkin, et al. The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, 2023. doi: 10.1093/genetics/iyad031.
- [63] M Kanehisa and S Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000. doi: 10.1093/nar/28.1.27.
- [64] M Kanehisa, M Furumichi, Y Sato, M Kawashima, and M Ishiguro-Watanabe. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51(D1):D587–D592, 2023. doi: 10.1093/nar/gkac963.
- [65] Y Benjamini and Y Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.