Mining the Krogan SARS-CoV-2 Human-Protein Interactome

A network analysis of the host machinery the virus engages, the physical complexes behind those signals, and a testable prediction for one missing pore component.

Analysis of Gordon et al., "A SARS-CoV-2 protein interaction map reveals targets for drug repurposing," Nature 2020 (DOI 10.1038/s41586-020-2286-9), retrieved from the EBI IntAct database.

---

1. The dataset

I retrieved the high-confidence affinity-purification mass-spectrometry (AP-MS) interactome directly from the EBI IntAct database, using the curated deposition behind the Krogan-lab paper and keyed on the paper's DOI. The retrieval returned exactly 332 interactions linking 26 SARS-CoV-2 viral proteins (baits) to 332 distinct human host proteins (prey), all assayed in HEK293T cells. These figures match the published map.

One structural feature is worth stating at the outset. In this high-confidence set, every host protein is targeted by exactly one viral protein, so the bipartite graph reduces to a disjoint union of 26 "stars," one per viral bait (Figure 1). The biologically interesting cross-talk, meaning the shared machinery the virus attacks, does not appear among shared baits. It appears in the prey-to-prey interactions, which is where the complex analysis in Section 3 focuses.

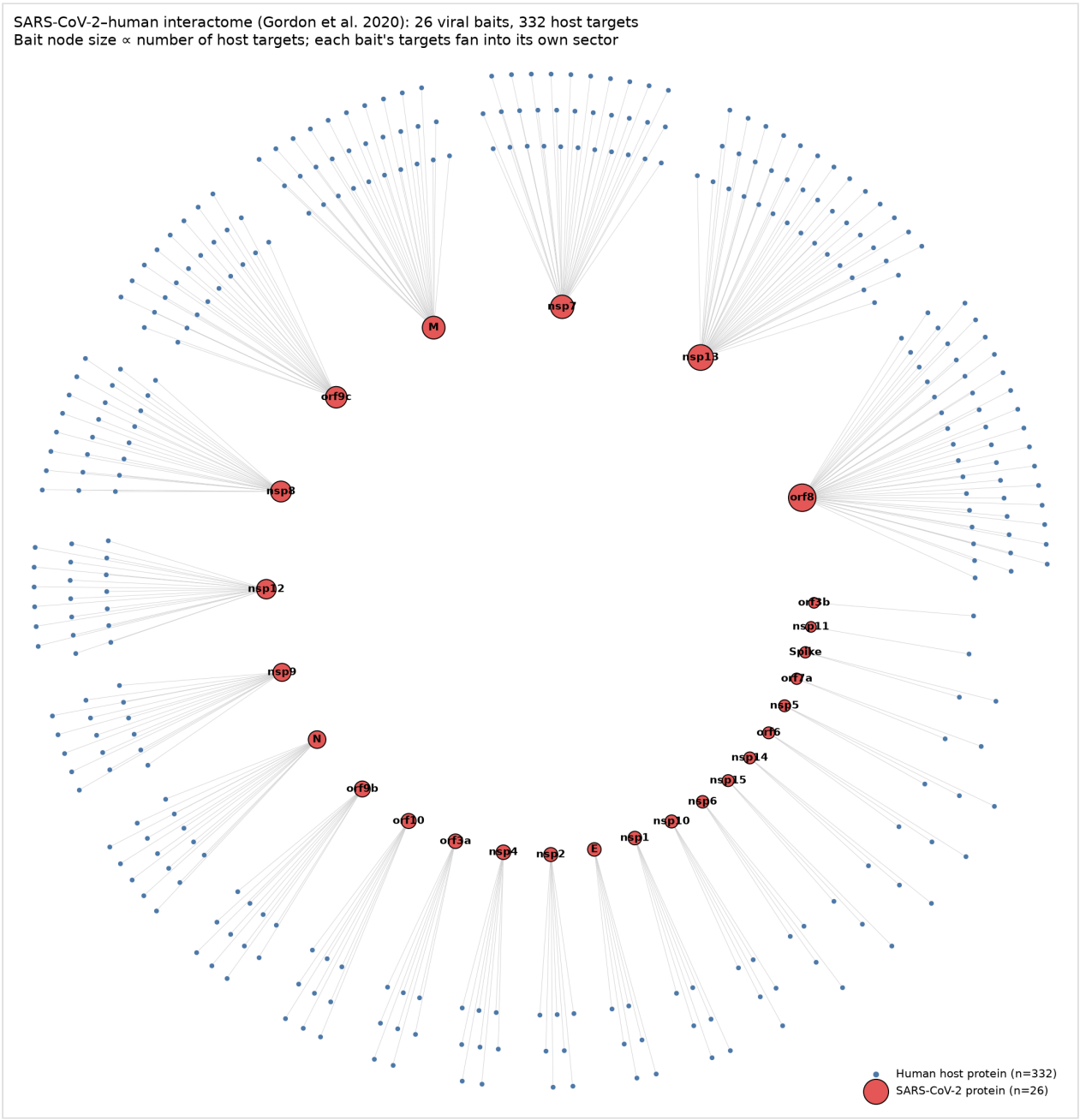

Figure 1. The SARS-CoV-2 human-protein interactome (Gordon et al., 2020): 26 viral baits and 332 host targets. Bait node size is proportional to the number of host targets, and each bait's prey fan into its own sector. Because every host protein is hit by a single viral protein, the graph forms 26 separate stars.

The target load is highly uneven (Figure 2). Three viral proteins together account for roughly 36 percent of all interactions:

Viral protein	Host targets	Role
orf8	47	Immune evasion / ER-lumen accessory
nsp13	40	Helicase
nsp7	32	RdRp cofactor
M	30	Membrane / assembly
orf9c	26	Accessory (membrane)

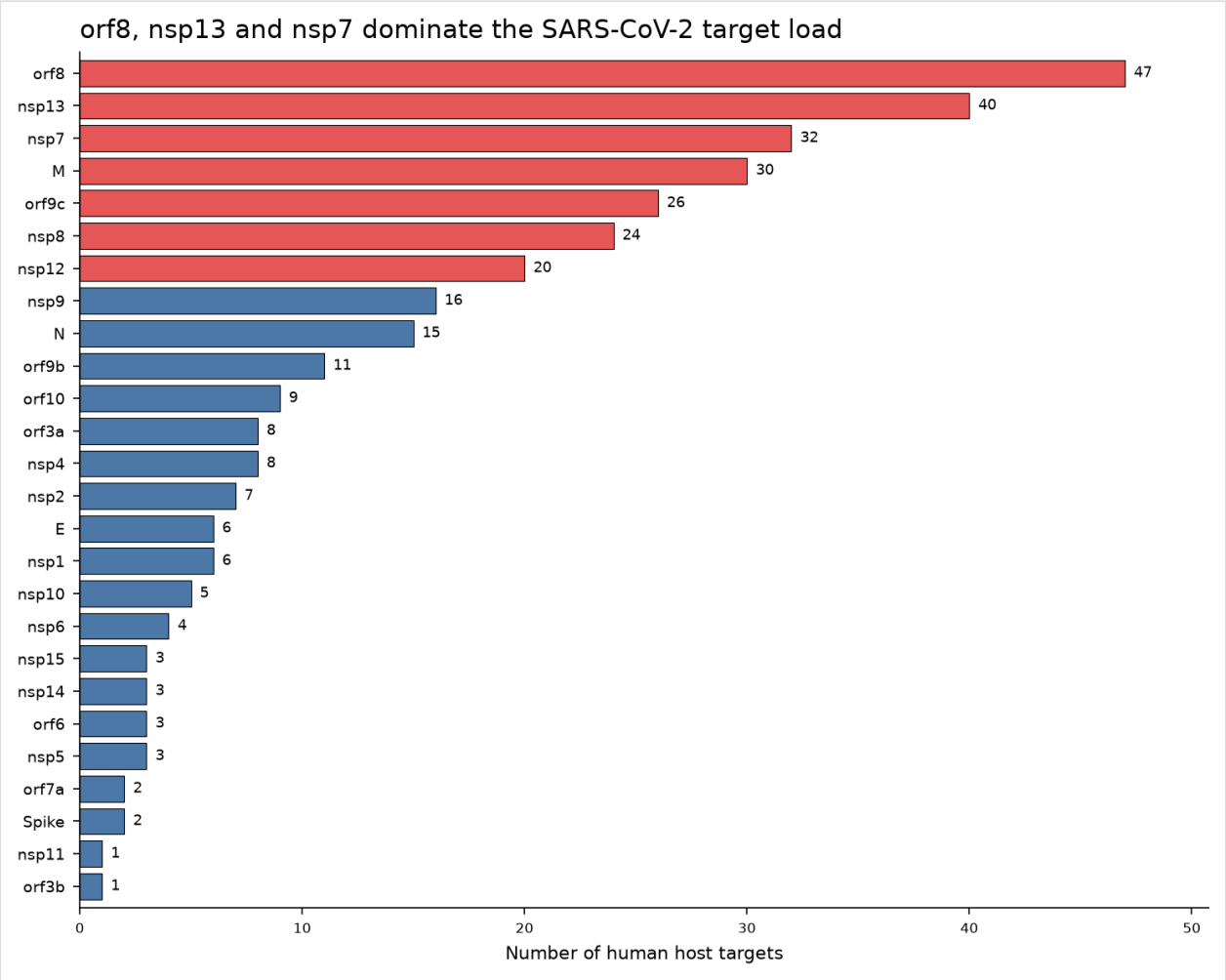

Figure 2. Host targets per viral protein. orf8, nsp13 and nsp7 carry the heaviest target loads, while most other viral proteins engage only a handful of host factors.

2. Which pathways are hijacked

STRING functional enrichment across all 332 prey returned 735 significantly enriched terms after FDR correction. The full table is provided as `pathway_enrichment.csv`, and Figure 3 shows a curated view of the leading terms.

The dominant themes, ranked by both significance and fold-enrichment, are the following.

- **Protein trafficking and localization** form the single most significant block (GO "Protein localization," FDR ≈ 3×10^{-13}; "Intracellular protein transport," 46 genes). The virus reaches broadly into how the cell routes proteins to membranes and organelles.
- **Mitochondrial protein import** is represented by 12 prey at 11-fold enrichment (Reactome, FDR ≈ 1×10^{-6}), covering the TIM/TOM import machinery and respiratory-chain assembly factors.
- **Nuclear pore complex disassembly** is 15-fold enriched (FDR ≈ 8×10^{-6}), reflecting perturbation of nucleocytoplasmic transport, a theme revisited in Section 4.
- **Mitotic spindle and centrosome** terms capture the recruitment of centrosome proteins and NuMA to spindle poles, with "M Phase" and "Cell Cycle, Mitotic" both strongly enriched.

- **Viral mRNA synthesis and ER protein processing** round out the specific machinery.

These themes are consistent with known coronavirus biology. SARS-CoV-2 rewires the secretory and trafficking system it needs to build replication organelles, co-opts mitochondrial and innate-immune machinery, and perturbs control of the cell cycle and nuclear transport.

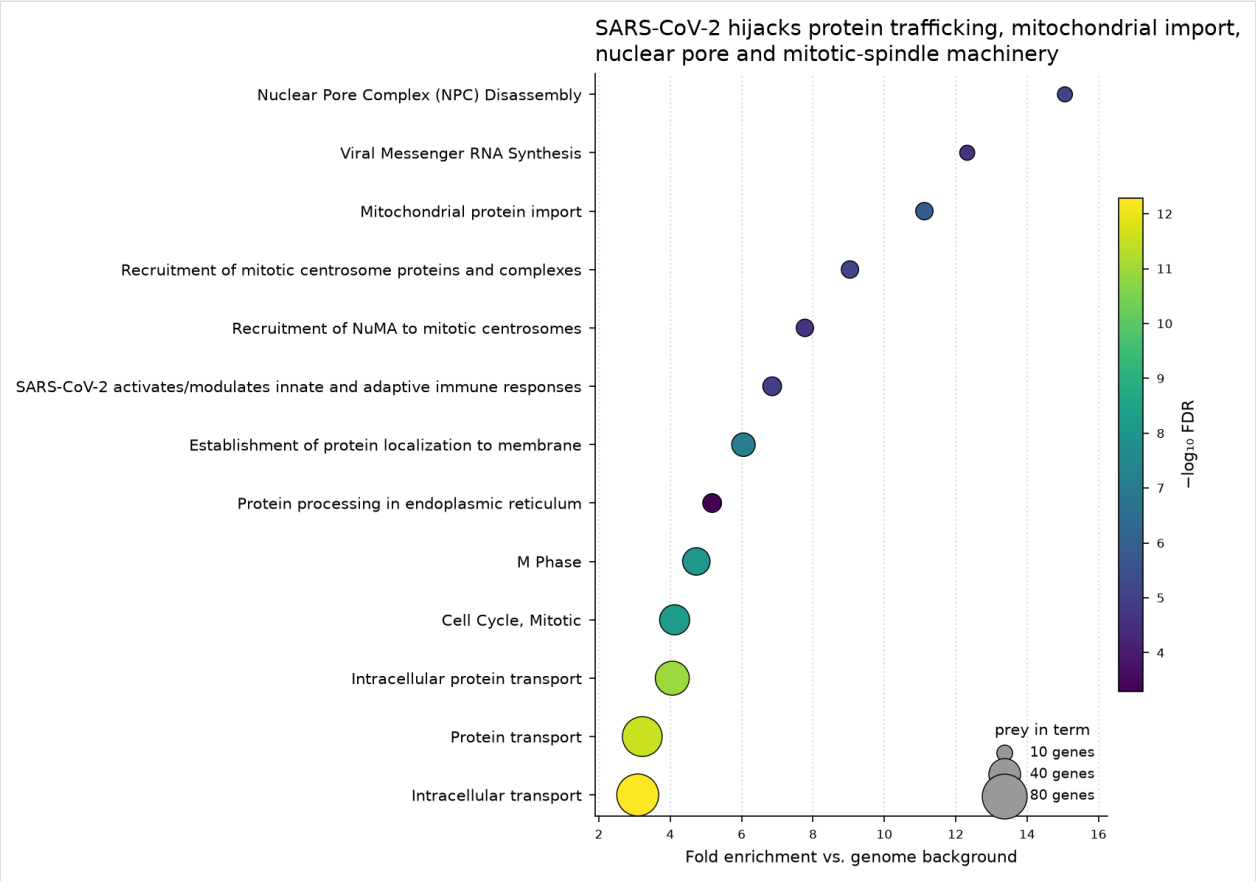

Figure 3. Curated STRING functional enrichment for the 332 host prey. Point position shows fold enrichment versus the genome background, color encodes significance, and point size reflects the number of prey in each term. Protein trafficking, mitochondrial import, nuclear pore and mitotic-spindle machinery lead the list.

3. Which complex is doing the work

Enrichment identifies processes rather than physical assemblies. To find the actual machines, I built the STRING high-confidence prey-to-prey interaction network (combined score ≥ 0.7), in which 224 of the 332 prey are interconnected, and applied Louvain community detection. This recovered 11 modules of at least five proteins. Each corresponds to a recognizable stable complex dominated by one or two viral proteins (Figure 4, `targeted_complexes.csv`).

Complex / machine	Prey	Dominant viral protein(s)	Enrichment FDR
Centrosome / mitotic spindle	28	nsp13	5×10^{-16}
Mitochondrial import & respiratory-chain assembly	27	orf9c, M, nsp4	2×10^{-18}
CUL2 ubiquitin-ligase & chromatin regulators	25	orf10	5×10^{-7}
Golgi / vesicle trafficking (Rab GTPases)	22	nsp7	4×10^{-11}

Complex / machine	Prey	Dominant viral protein(s)	Enrichment FDR
ER protein processing / SRP	17	orf8, nsp8	2×10^{-8}
Endocytosis / small-GTPase signaling	14	nsp7, M	7×10^{-7}
RNA exosome & ribosome biogenesis	12	nsp8	8×10^{-13}
Stress granule / mRNP	12	N	3×10^{-16}
MARK kinases & innate-immune signaling	9	orf9b	7×10^{-4}
Nuclear pore complex	9	nsp9	3×10^{-18}
Cytochrome / redox	5	nsp7	3×10^{-7}

The central result is that the interactome resolves into a small number of intact host complexes, each engaged by a specific viral protein. The nucleocapsid (N) protein converges on the stress-granule and mRNP system, including the antiviral stress-granule hub G3BP1/G3BP2. orf9c and M attack mitochondrial import, nsp8 targets the RNA exosome and ribosome biogenesis, and nsp9 binds the nuclear pore complex as a tight, well-separated module.

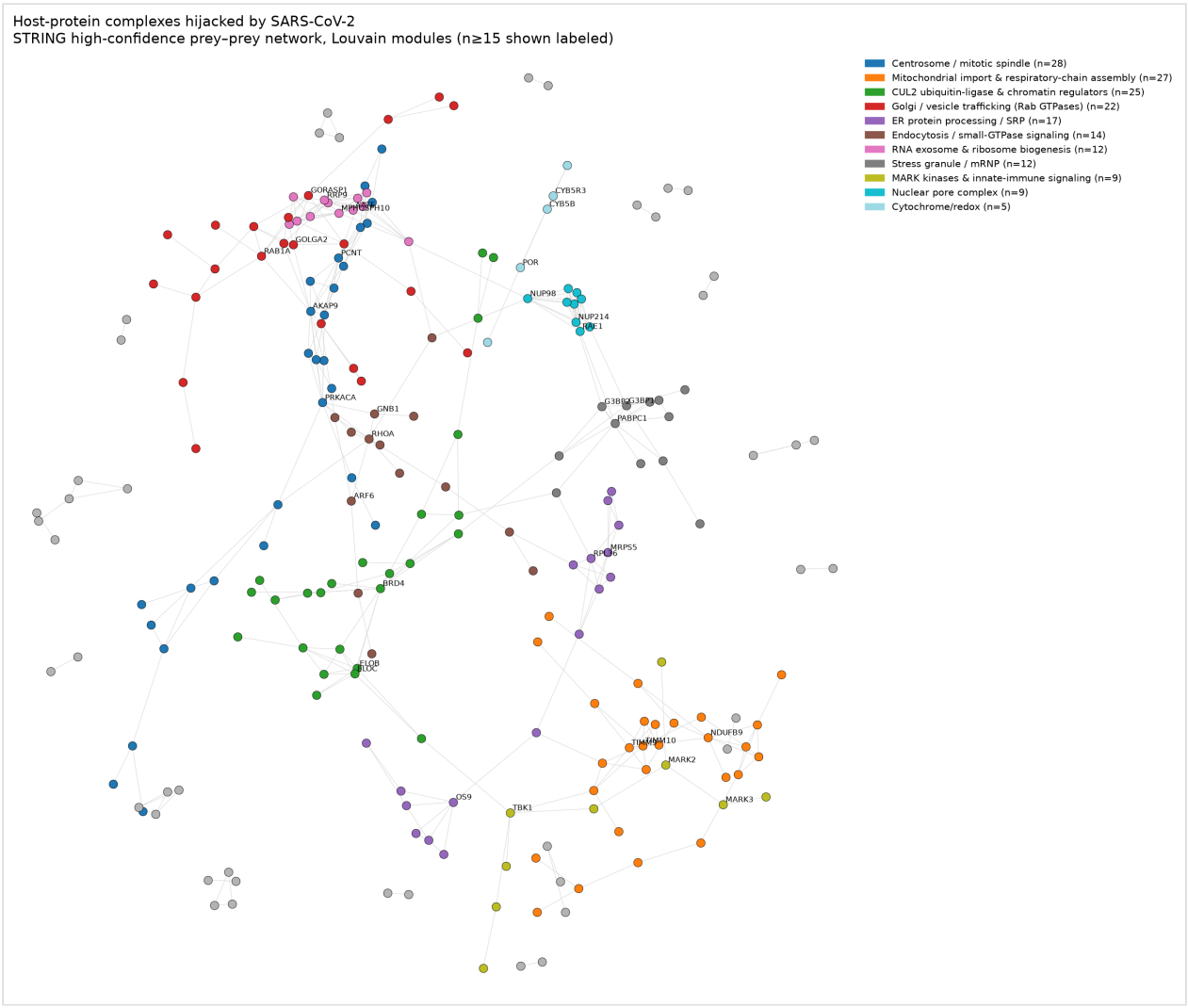

Figure 4. Host-protein complexes hijacked by SARS-CoV-2. The STRING high-confidence prey-to-prey network is colored by Louvain module, with the larger modules labeled. Each module maps onto a recognizable stable complex dominated by one or two viral proteins.

4. A predicted missing component: NUP93

I selected the nuclear pore complex (NPC) as the focal complex for prediction because it is a well-defined, structurally characterized machine, and because Krogan captured only a partial subset of it through nsp9.

What Krogan captured. Cross-referenced against the curated NPC record in Complex Portal (CPX-873, 32 canonical members), the Krogan prey in this module include eight canonical structural nucleoporins, namely NUP62, NUP54, NUP58, NUP88, NUP98, NUP210, NUP214 and RAE1, together with the transport factor NUTF2. This set contains the complete central-channel FG-nucleoporin trimer NUP62/NUP54/NUP58, which forms the permeability barrier of the pore.

The missing component and the prediction. Of the 24 canonical NPC members not captured, I ranked candidates on two criteria: STRING connectivity to the eight captured structural nucleoporins, and specific anchoring to the captured central-channel trimer. The top-ranked candidate is NUP93 (`predicted_missing_component.csv`, Figure 5). NUP93 connects to all eight captured structural nucleoporins at approximately 0.99 STRING confidence, and it links to the channel trimer NUP62, NUP54 and NUP58 at 0.999, 0.999 and 0.999 respectively.

Biological rationale. This ranking is not an arbitrary connectivity artifact. The NUP93 N-terminal domain is the physical adaptor that tethers the NUP62/NUP54/NUP58 channel trimer to the NPC inner-ring scaffold. The Krogan AP-MS experiment recovered the channel trimer but not its anchor, NUP93. A topology-based reading of the map therefore predicts NUP93 as the host factor most likely to complete the hijacked pore module, whether through an interaction that fell below the AP-MS detection threshold or one bridged indirectly through the captured trimer.

How to test it. A targeted co-immunoprecipitation of nsp9 probing for NUP93, together with the reciprocal NUP93 pulldown probing for nsp9, or a proximity-labeling experiment (BioID or APEX) using nsp9 as bait, would confirm or refute the prediction directly. A negative result would be informative in its own right, indicating that nsp9 engages the channel trimer independently of the inner-ring anchor.

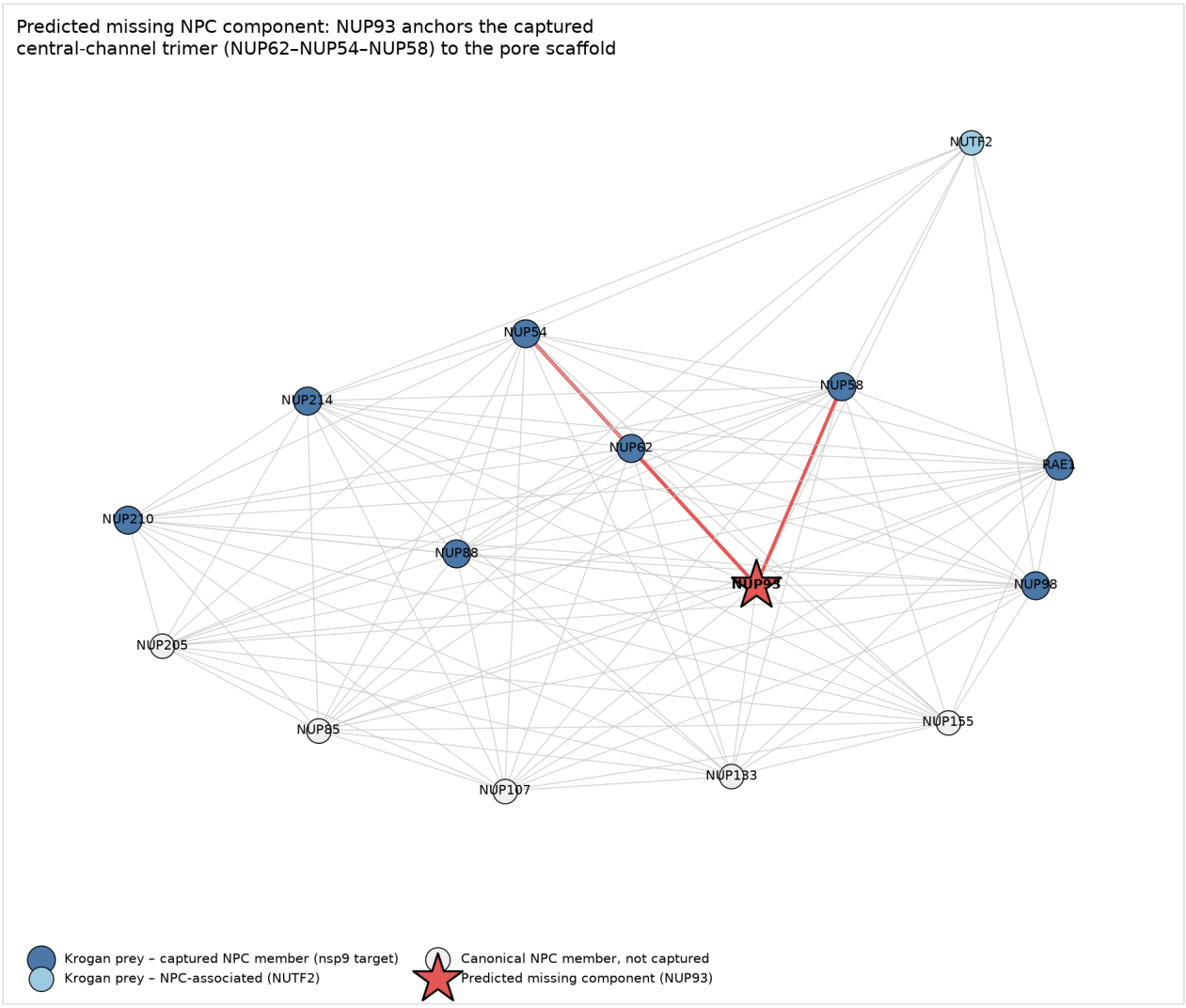

Figure 5. Predicted missing NPC component. Within the nucleoporin subnetwork, NUP93 (star) is the highest-ranked canonical member absent from the Krogan map. Its N-terminal domain anchors the captured central-channel trimer NUP62/NUP54/NUP58 (highlighted links) to the pore scaffold.

5. Caveats

- AP-MS captures proximity and co-complex membership rather than direct binary binding. A viral protein that appears to target a complex may bind one member directly and co-purify the remainder.

- The map uses the spoke-expansion model, connecting each bait to each of its prey, so the absence of shared baits noted in Section 1 is partly a property of how the data are represented rather than purely a feature of the underlying biology.
- The prey-to-prey network and all enrichment analyses draw on STRING, which combines physical evidence with functional and predicted evidence. I applied the high-confidence threshold (≥ 0.7) for the network to favor physical complexes.
- The NUP93 nomination is a computational hypothesis derived from network topology and known NPC architecture, not an experimental finding. It is offered as a testable prediction rather than a claim.

Artifacts

File	Contents
`krogan_sars2_ppi.csv`	The 332 viral-to-host interactions, with readable viral labels
`sars2_interactome_network.png`	Bipartite interactome, baits sized by target load
`prey_per_viral_protein.png`	Host targets per viral protein
`pathway_enrichment.csv`	Full STRING enrichment (735 terms)
`enriched_pathways.png`	Curated enriched-pathway dot plot
`targeted_complexes.csv`	11 host complexes, members, and targeting viral proteins
`complex_modules.png`	Module-colored prey-to-prey complex network
`predicted_missing_component.csv`	Ranked NPC missing-component candidates
`missing_component_prediction.png`	NUP93 prediction within the NPC subnetwork