Mapping & Querying Protein-Protein Interactions from SARS-CoV-2, HPV, and HIV Datasets

Shanzay Farzan
Middlesex County College

Advisor: Dr. Phalguni Ghosh

Respiratory Droplet
2020, Watercolor and Ink on Paper

RCSB Protein Data Bank
doi: 10.2210/rcsb_pdb/goodsell-gallery-024

SARS-CoV-2 PDB Timeline

Late 2019
- Unexplained influenza-like illness seen in Wuhan, PRC

January 2020
- Identified as SARS-CoV-2 virus
- Genome sequence released

May 2021
- Worldwide cases: 154 Million
- Number of cases in the US: 32.5 Million
- Number of deaths in the US: 578,000

"A SARS-CoV-2 protein interaction map reveals targets for drug repurposing"

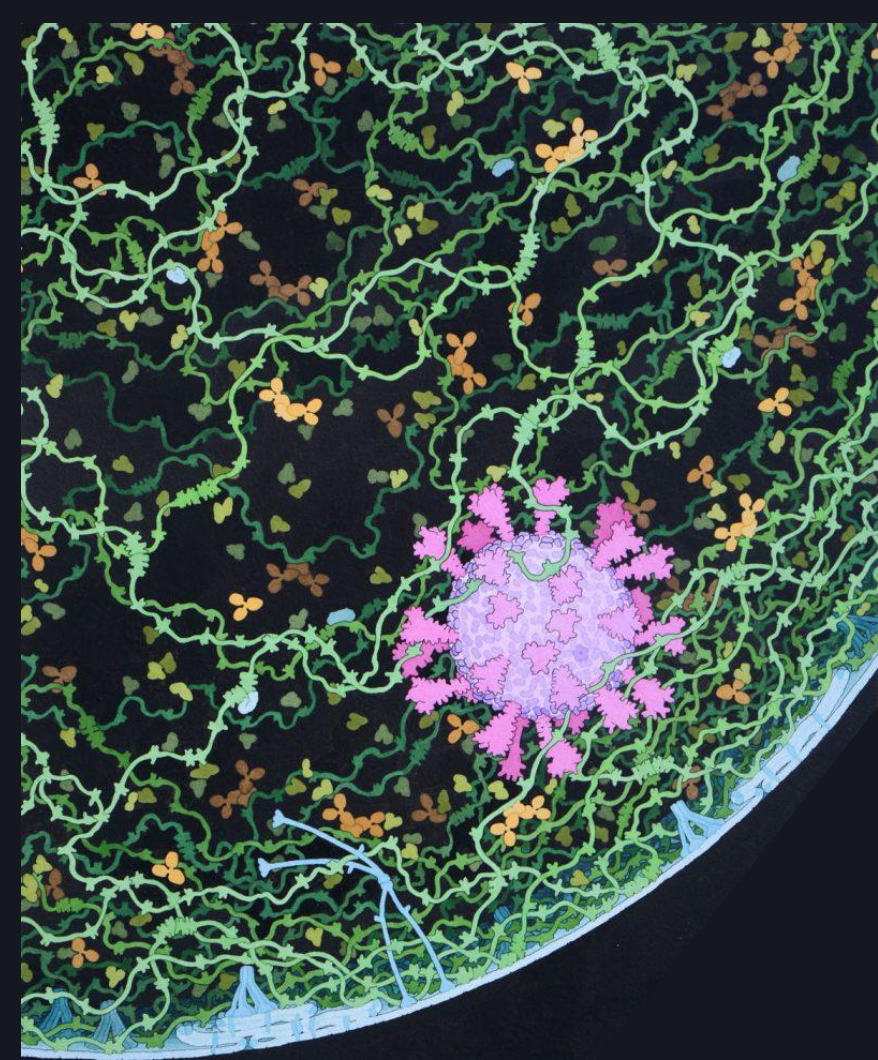Gordon, D.E., Jang, G.M., Bouhaddou, M. *et al. Nature* 583**,** p. 459-468 (2020).

1. Identified human proteins that interact with SARS-CoV-2 protein
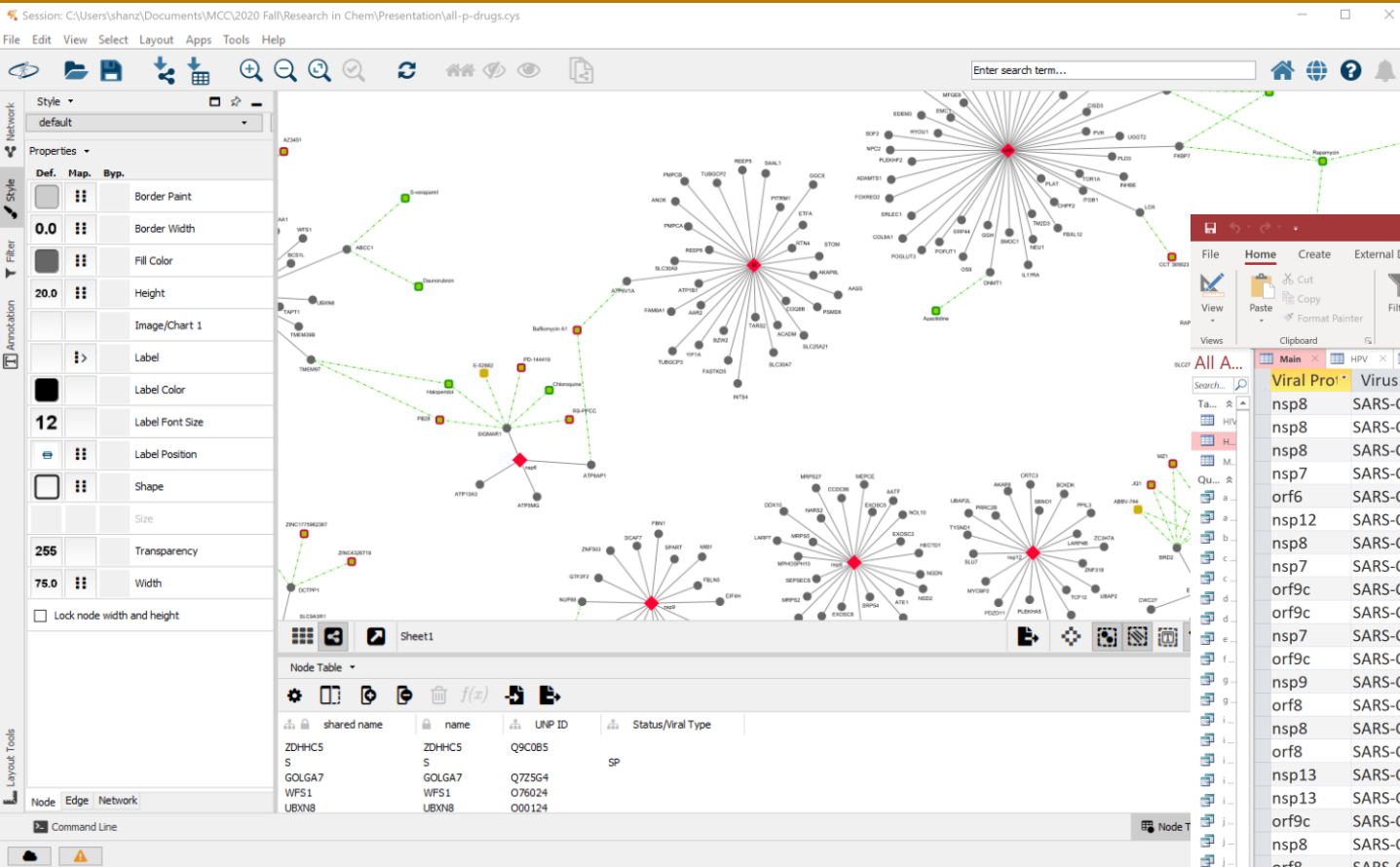2. Identified 66 potential human proteins targeted by 69 compounds

Objectives

1. Study **SARS-CoV-2 Human-Protein (HP) interactions** through **visual networks** and extend to other viral pathogens

2. Create **composite SARS-CoV-2 network** showing compounds that can inhibit viral/human protein interactions and extend to other viral pathogens

3. Build a **database** of viral/human proteins and compounds that inhibit these interactions

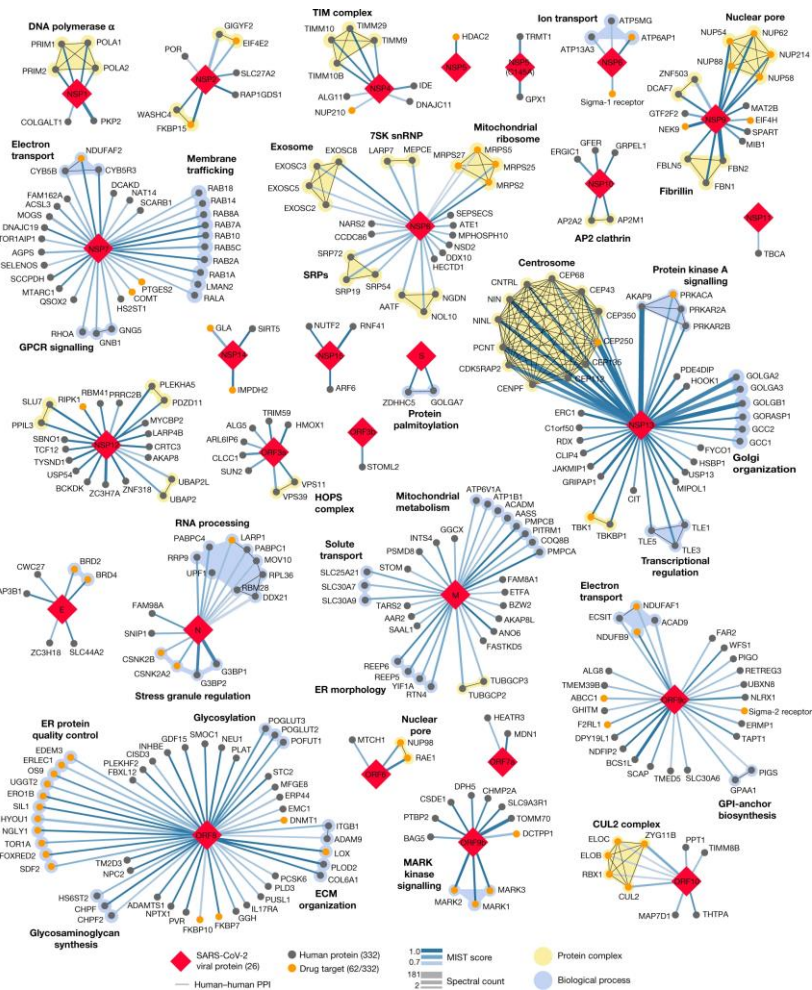4. Query the database using **SQL**

Software Used

MS Access

Cytoscape

Research Articles and Networks

"A SARS-CoV-2 protein interaction map reveals targets for drug repurposing"

Gordon, D.E., Jang, G.M., Bouhaddou, M. *et al.*
Nature 583**,** p. 462 (2020).

"Global landscape of HIV-human protein complexes"

Jäger, S., Cimermancic, P., Gulbahce, N. *et al.*
Nature 481**,** p. 368 (2012).
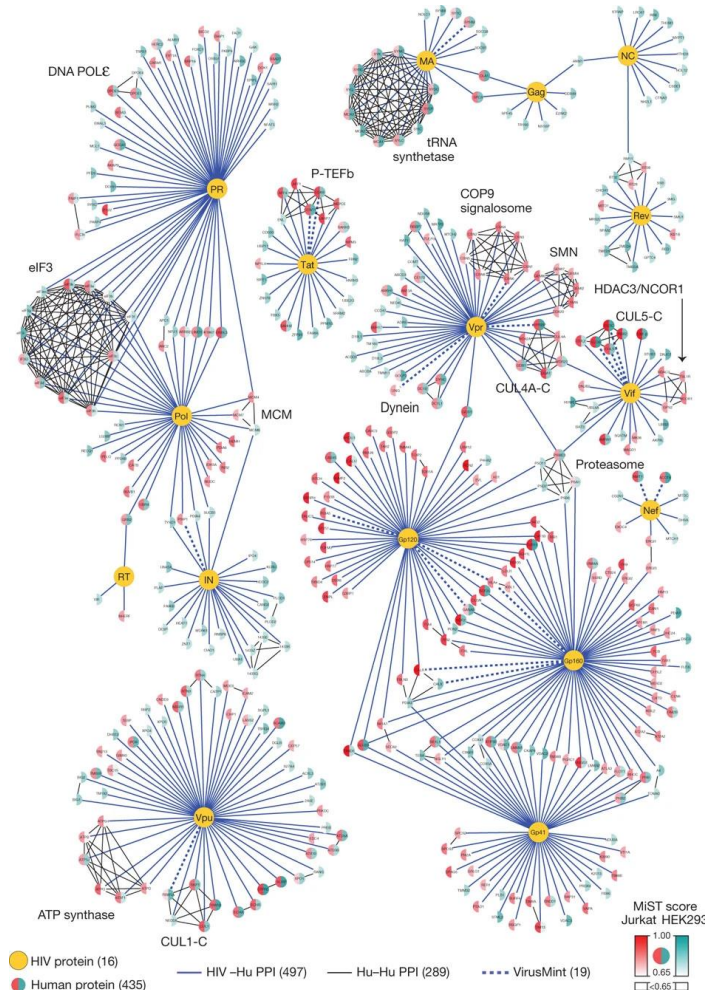
"Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus-Host Protein Network"

Eckhardt, M., Zhang, W., Gross, A.M. *et al.*
Cancer Discovery, American Association for
Cancer Research. Vol. 8, No. 11, p. 1478 (2018).

Source for all networks: **ndexbio.org**

Viral Life Cycle

Capsid

RNA

Envelope (with glycoproteins)

1 Glycoproteins on the viral envelope bind to specific receptor molecules (not shown) on the host cell, promoting viral entry into the cell.

2 The capsid and viral genome enter the cell. Digestion of the capsid by cellular enzymes releases the viral genome.

HOST CELL

Viral genome (RNA)

Template

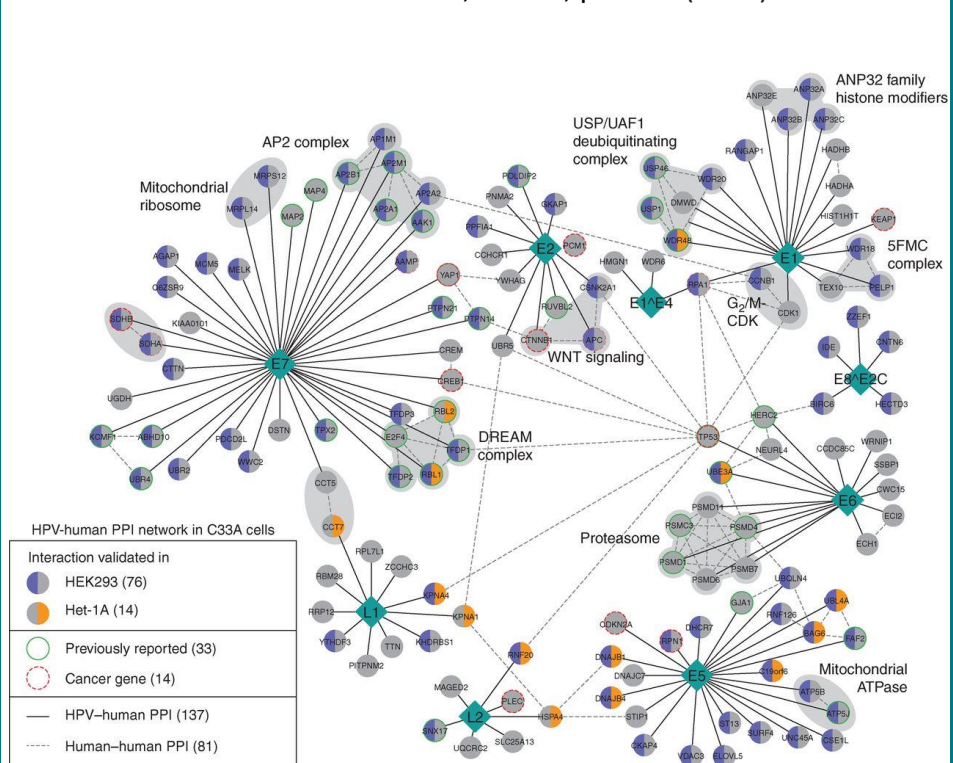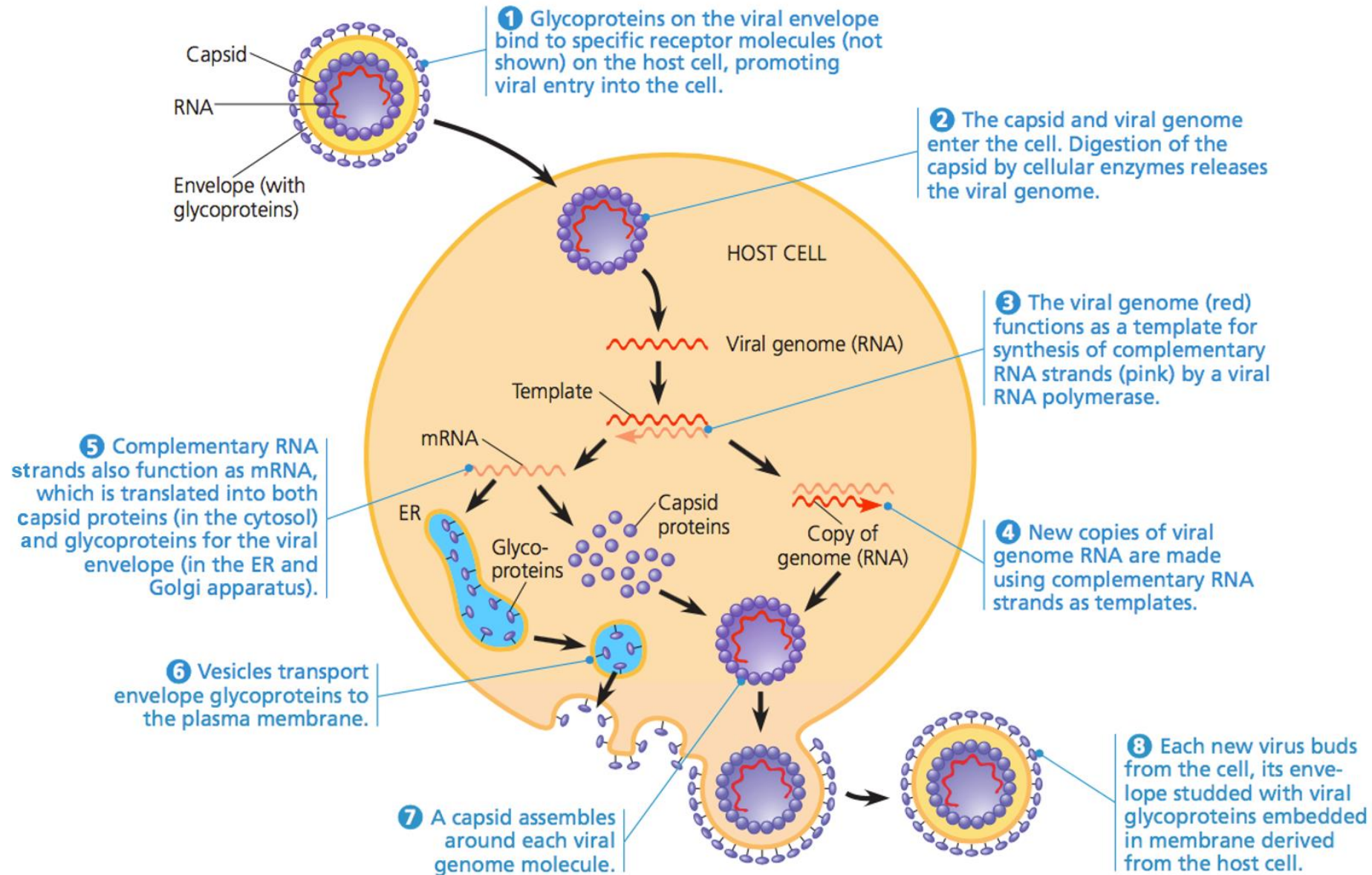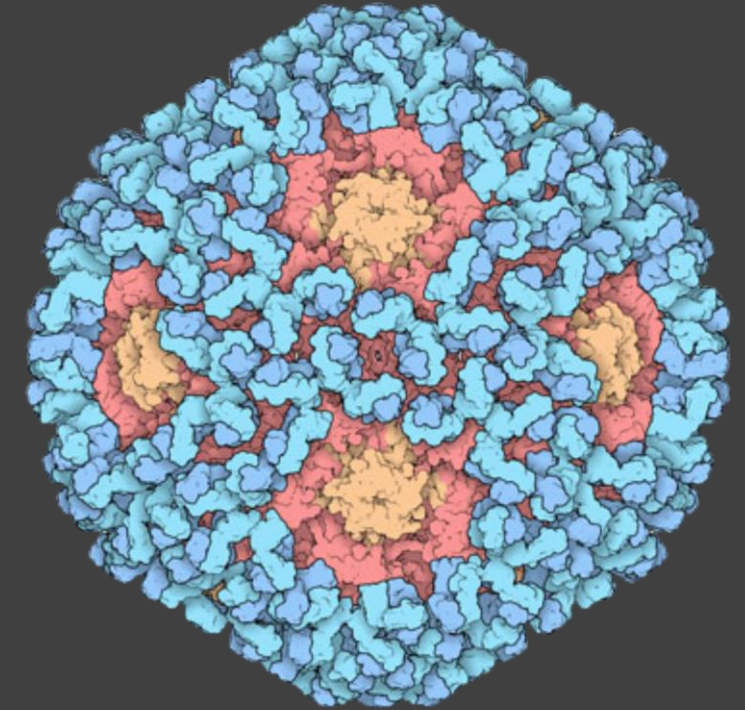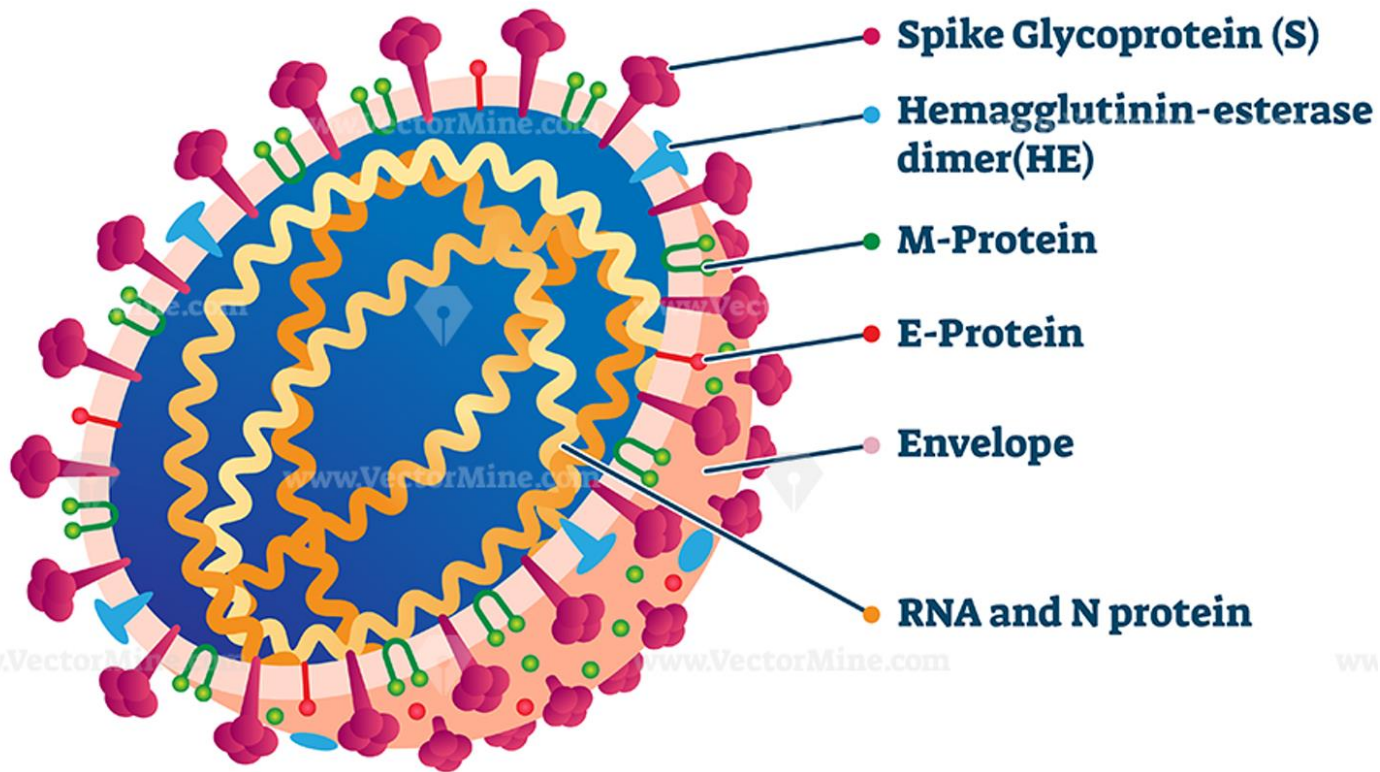**3** The viral genome (red) functions as a template for synthesis of complementary RNA strands (pink) by a viral RNA polymerase.

mRNA

ER

5 Complementary RNA strands also function as mRNA, which is translated into both capsid proteins (in the cytosol) and glycoproteins for the viral envelope (in the ER and Golgi apparatus).

Glycoproteins

Capsid proteins

Copy of genome (RNA)

4 New copies of viral genome RNA are made using complementary RNA strands as templates.

6 Vesicles transport envelope glycoproteins to the plasma membrane.

7 A capsid assembles around each viral genome molecule.

8 Each new virus buds from the cell, its envelope studded with viral glycoproteins embedded in membrane derived from the host cell.

Human SARS-CoV-2

Human Immunodeficiency Virus

Human Papilloma Virus

Source: Protein Data Bank https://pdb101.rcsb.org/

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

Spike Glycoprotein (S)

Hemagglutinin-esterase dimer(HE)

M-Protein

E-Protein

Envelope

RNA and N protein

Four structural proteins:
- Spike (S) – allows virus to penetrate the cell
- Envelope (E) – helps assemble new viruses
- Membrane (M) – viral assembly
- Nucleocapsid (N) – allows the virus to camouflage its genetic material to the immune system

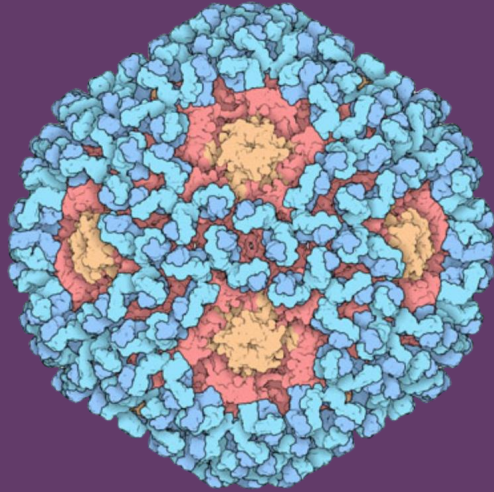**Sixteen Non-structural proteins:**
- Nsp1-16

Process:
- The Spike protein binds to ACE2 protein receptor site in human cell
- After fusing and entering the cytoplasm, the virus commandeers the ribosome and other protein-making machinery to make new copies of itself
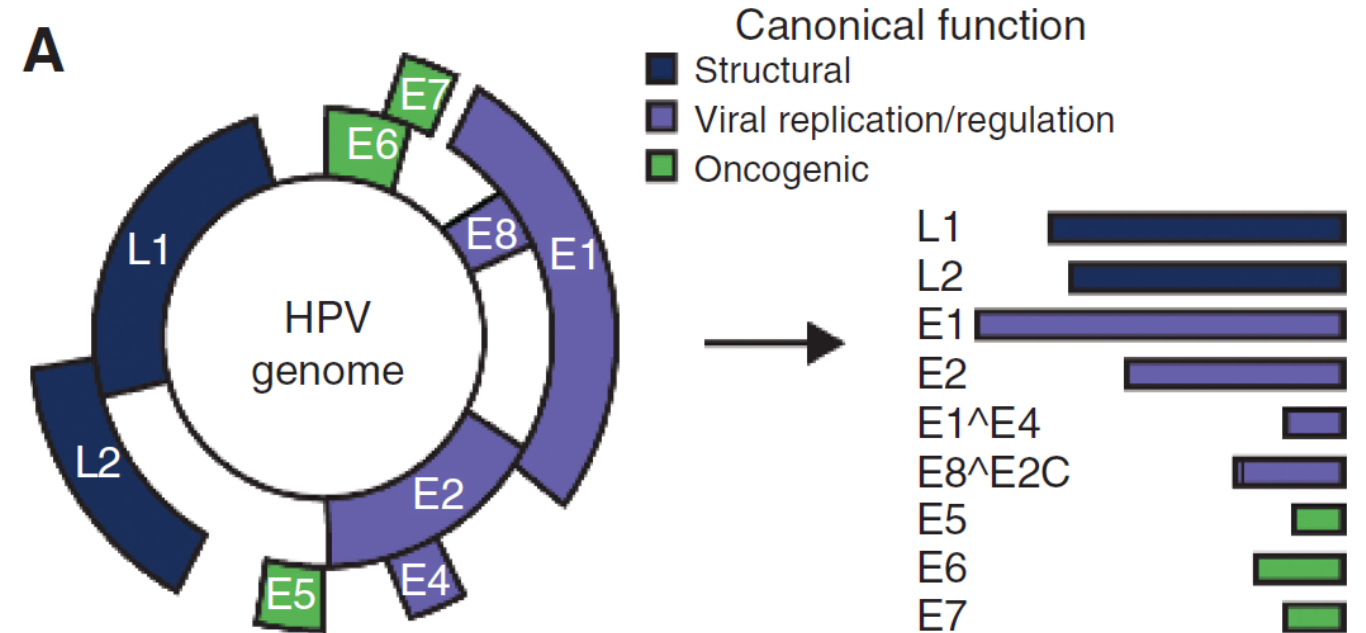
HPV Structure

- The capsid includes 360 copies of the major capsid chain, called L1. A second capsid chain, called L2, is found on the inside and may help with packaging the genome.

- Cancer: The genome encodes two small proteins, called E6 and E7, that act as cancer-causing oncogenes in dangerous strains. They stimulate unnatural growth of cells and block their natural defenses.
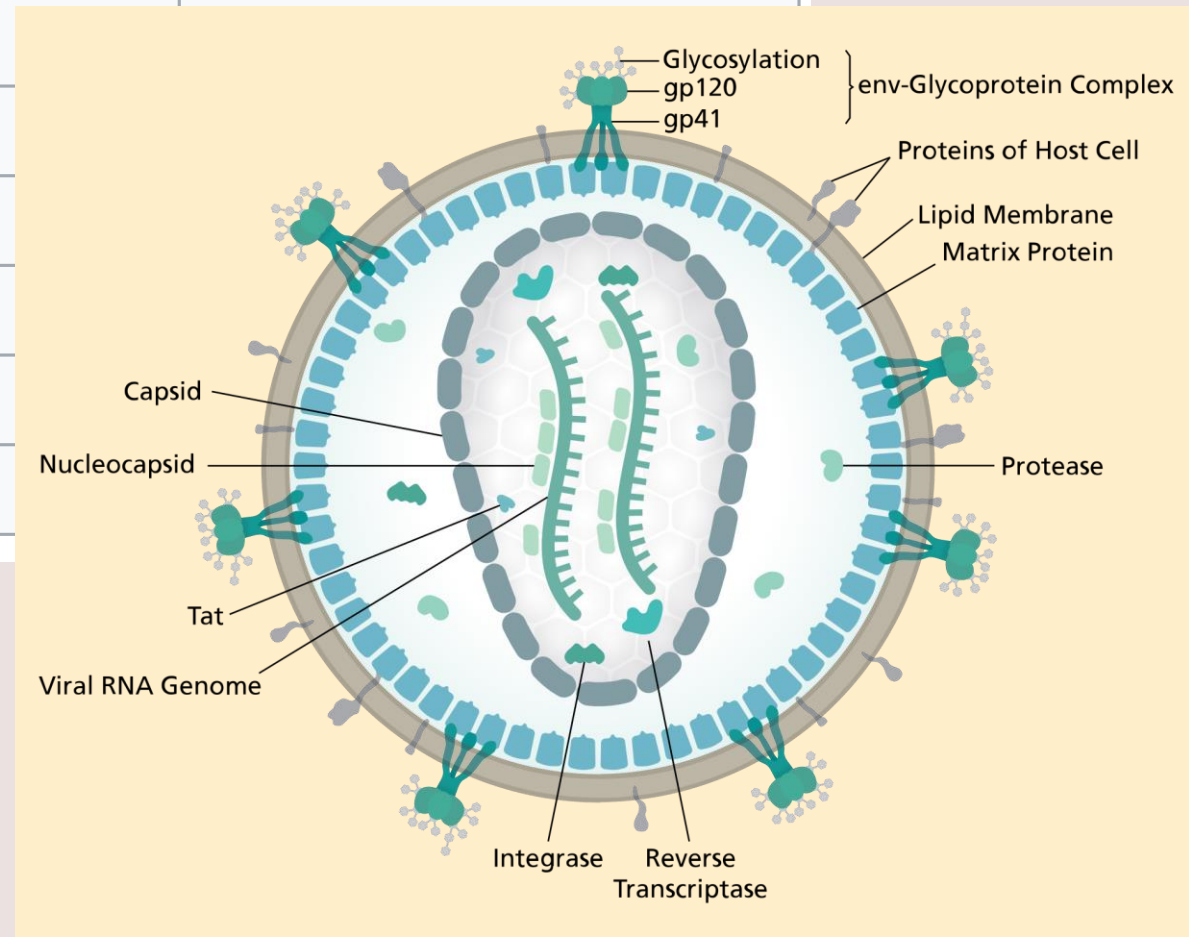
https://pdb101.rcsb.org/motm/221

HPV–Host Network Promotes Multiple Routes to Oncogenesis

A

Canonical function
- Structural
- Viral replication/regulation
- Oncogenic

HPV genome

L1
L2
E1
E2
E1^E4
E8^E2C
E5
E6
E7

Proteins encoded by the HIV genome

Class	Gene name	Primary protein products	Processed protein products
Viral structural proteins	*gag*	Gag polyprotein	MA, CA, SP1, NC, SP2, P6
	pol	Pol polyprotein	RT, RNase H, IN, PR
	env	gp160	gp120, gp41
Essential regulatory elements	*tat*	Tat	
	rev	Rev	
Accessory regulatory proteins	*nef*	Nef	
	vpr	Vpr	
	vif	Vif	
	vpu	Vpu	

HIV Structure

Labels: Glycosylation, gp120, gp41, env-Glycoprotein Complex, Proteins of Host Cell, Lipid Membrane, Matrix Protein, Protease, Reverse Transcriptase, Integrase, Viral RNA Genome, Tat, Nucleocapsid, Capsid

Li G, Piampongsant S et. al. "An integrated map of HIV genome-wide variation from a population perspective". *Retrovirology*. **12** (1): 18.
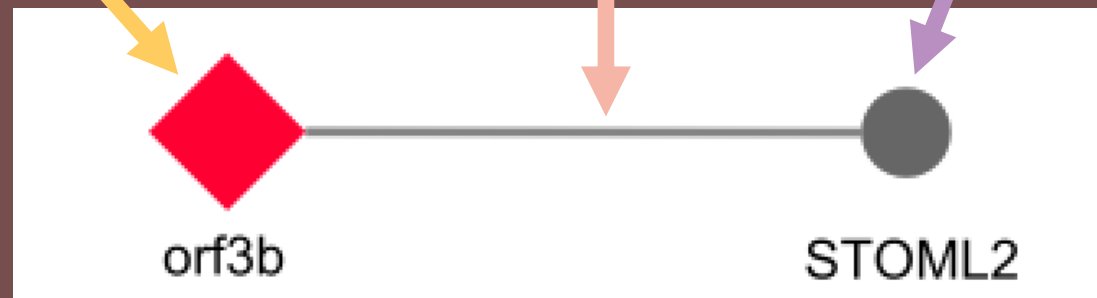
Create **visual networks** in Cytoscape showing interactions between viral and human proteins

| Viral Prot⁎	Virus Name	Viral Prot⁎	MIST	Human Gene	UNP ID	
orf3b	SARS-CoV-2	Non SP	0.929039041	STOML2	Q9UJZ1	stomatin like 2

Target node

Edge attribute

Source node

orf3b ——————— STOML2

Cytoscape

"Multiple Routes to Oncogenesis are Promoted by the Human Papillomavirus-Host Protein Network

Eckhardt, M., Zhang, W., Gross, A.M. *et al. Cancer Discovery*, American Association for Cancer Research. Vol. 8, No. 11, p. 1478 (2018).

Recreated Network in Cytoscape with MIST
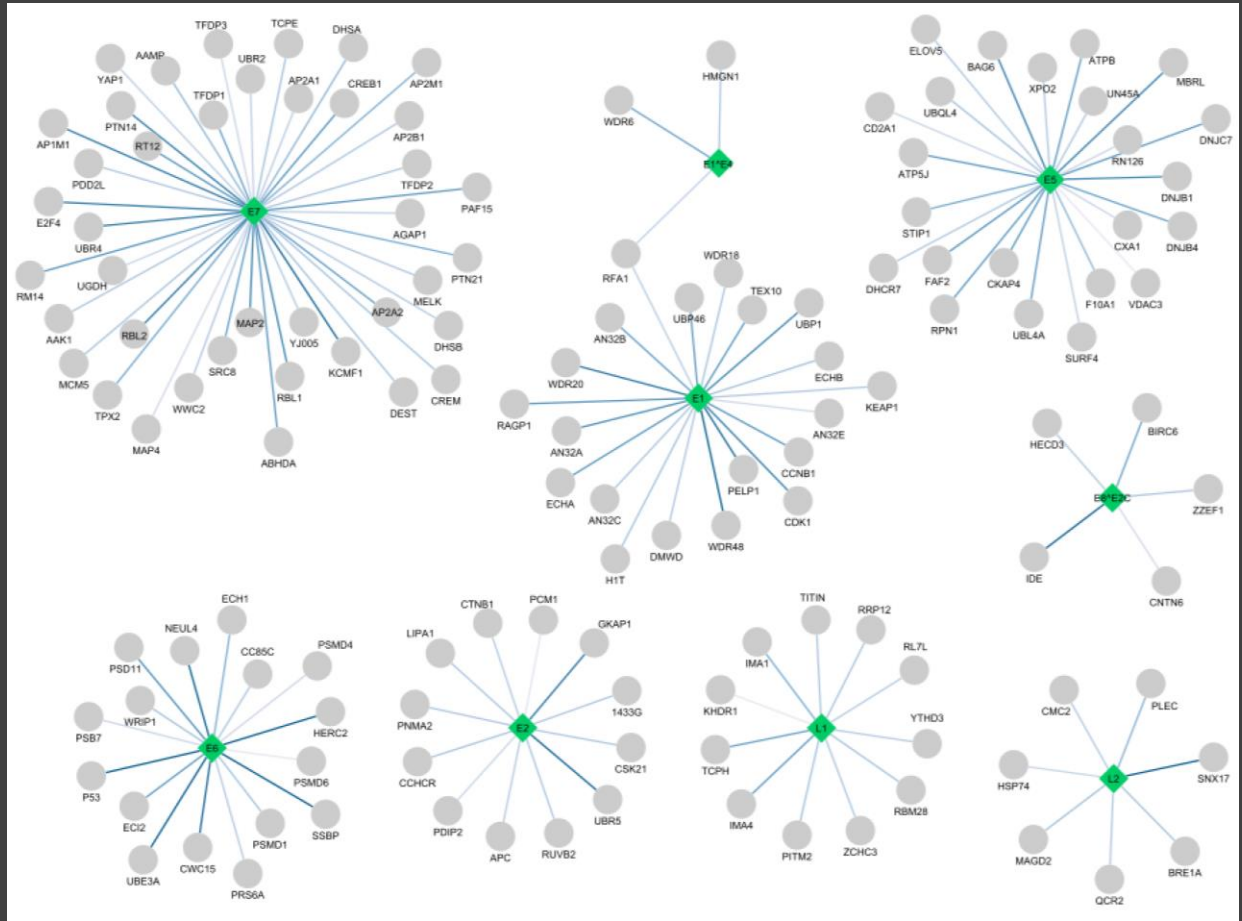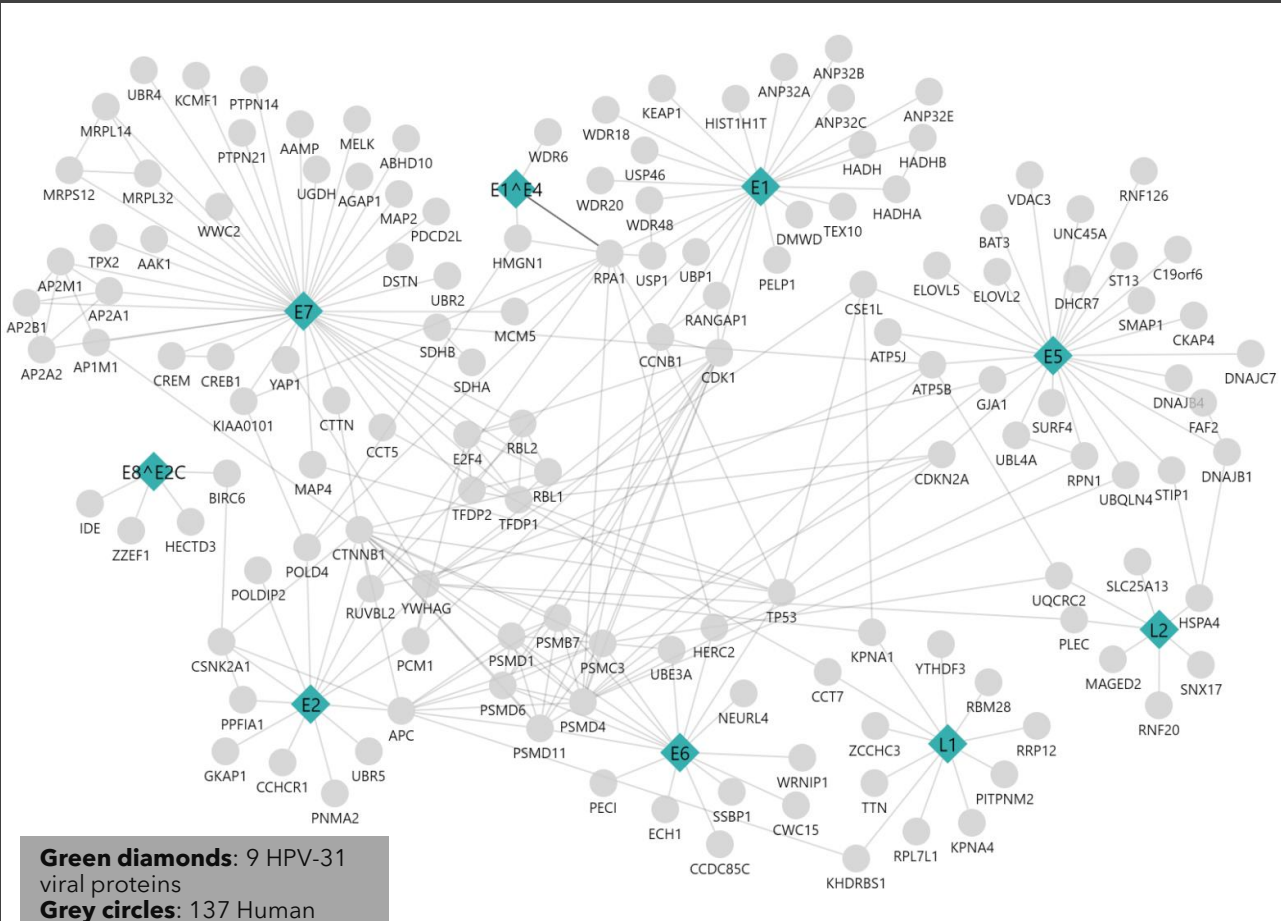
Green diamonds: 9 HPV-31 viral proteins
Grey circles: 137 Human Proteins
Grey edges: HPV-human interactions & human-human interactions.

Fig. 2: HPV-human Protein Network Map

MIST: Mass Spectrometry Interaction Statistics

- Weighted sum of three measures:
  - **Abundance:** Protein abundance measured by peak intensities from the mass spectrum
  - **Reproducibility:** invariability of abundance over replicated experiments
  - **Specificity:** Uniqueness of an observed host-pathogen interaction across all viral purifications

- For our purposes: it is a quantitative score to establish how strong protein-protein interactions are between the human gene and the viral protein

- Number from 0-1 and is always less than 1, where 0 represents weak interactions and numbers close to 1 represent high confidence interactions.

"A SARS-CoV-2 protein interaction map reveals targets for drug repurposing"

Gordon, D.E., Jang, G.M., Bouhaddou, M. et al. Nature 583, 459–468 (2020)
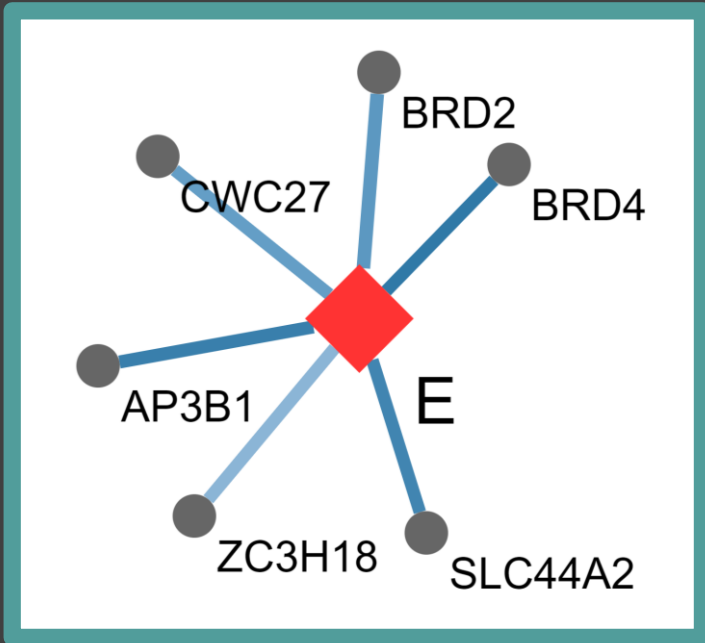
Red diamonds: SARS-CoV-2 bait proteins
Grey circles: human protein interaction partners
Blue edges: PPIs from this study. Darker blue indicates higher PPI confidence, thicker line indicates greater protein abundance in the AP-MS sample
Grey edges = PPIs from the CORUM

Fig. 3: SARS-CoV-2 Host-Pathogen Interaction Map
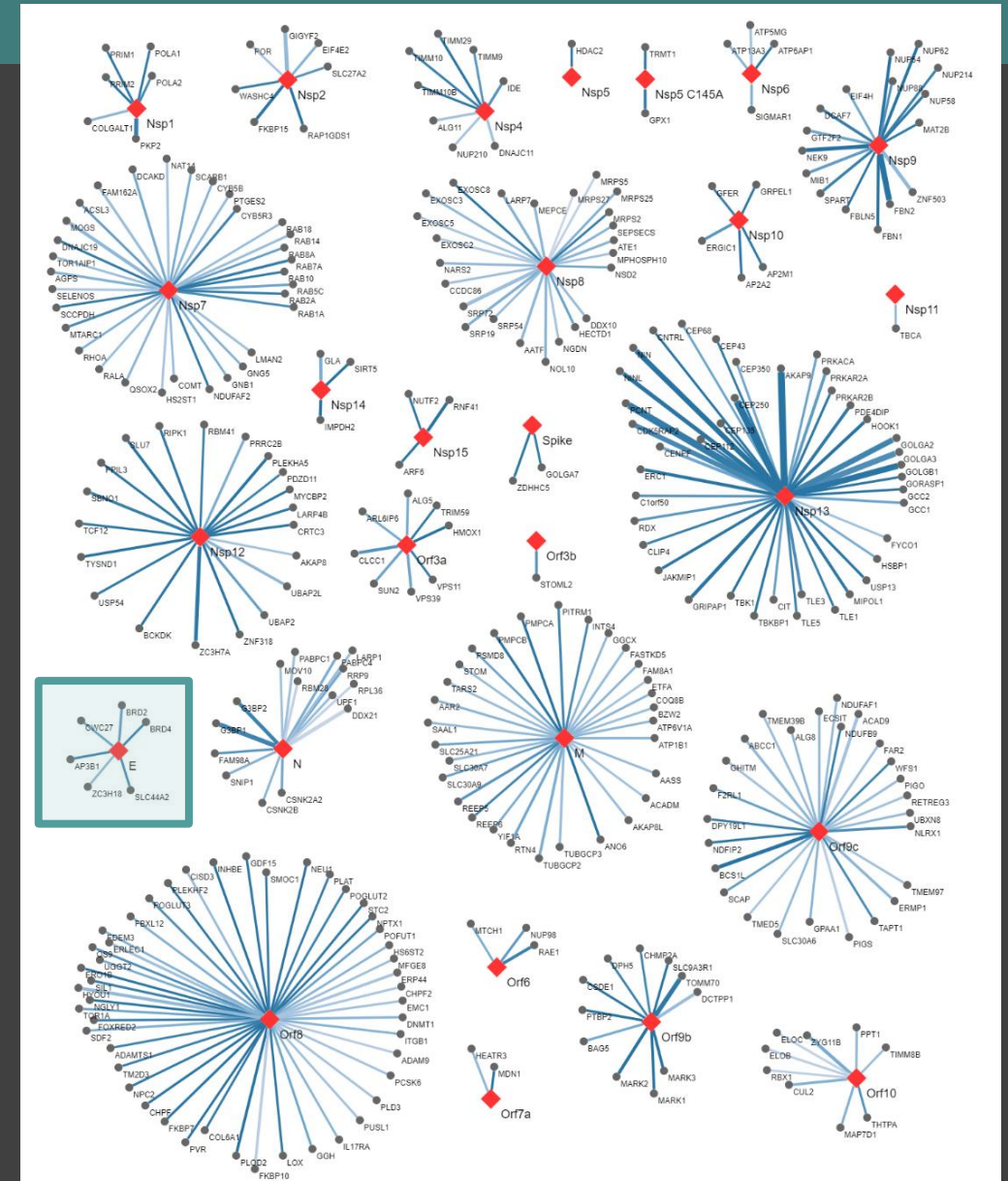
VIRAL PROTEIN	VIRUS NAME	VIRAL PROTEIN TYPE	MIST	HUMAN PROTEIN	UNP ID
E	SARS-CoV-2	SP	0.950342834	SLC44A2	Q8IWA5
E	SARS-CoV-2	SP	0.796415039	ZC3H18	Q86VM9
E	SARS-CoV-2	SP	0.89310916	CWC27	Q6UX04
E	SARS-CoV-2	SP	0.906592876	BRD2	P25440
E	SARS-CoV-2	SP	0.97848835	BRD4	O60885
E	SARS-CoV-2	SP	0.963550095	AP3B1	O00203

Literature-derived drugs and reagents that modulate SARS-CoV-2 interactors

COMPOUND NAME	HUMAN GENE	VIRAL BAIT	DRUG STATUS	ACTIVITY DESCRIPTION (NM)
JQ1	BRD2/4	E	Pre-clinical	BRD inhibitor, IC_{50} = 40-120
RVX-208	BRD2/4	E	Clinical Trial	BRD inhibitor, IC_{50} = 50-1800
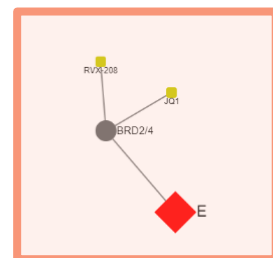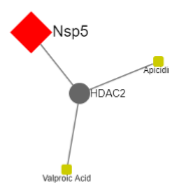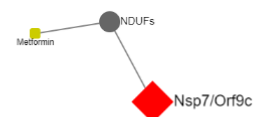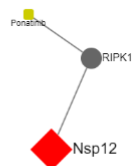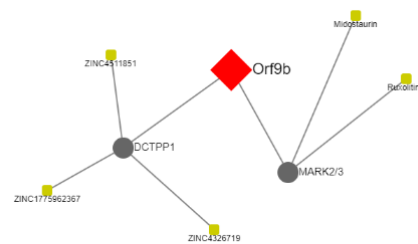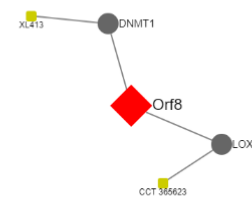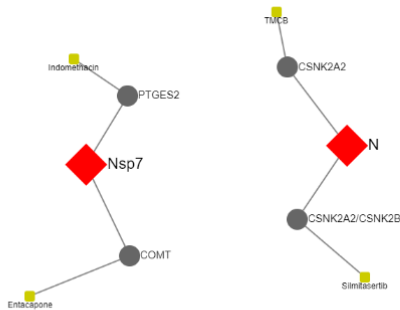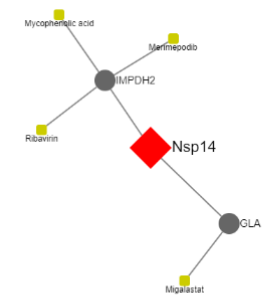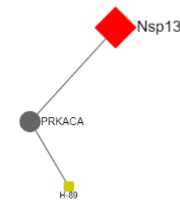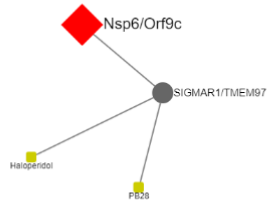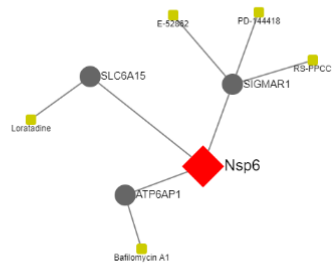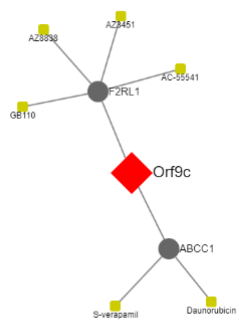
Supplementary Table 3 - Drug-target associations drawn from chemoinformatic searches of the literature, including information about purchaseability.
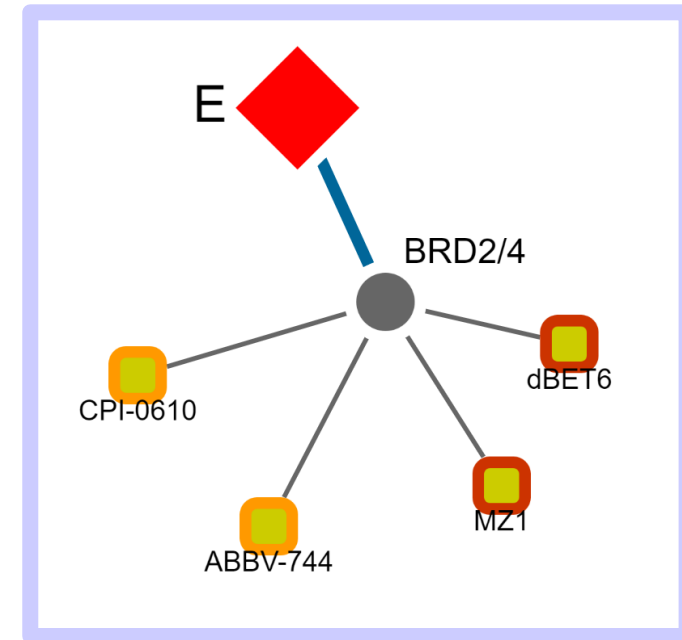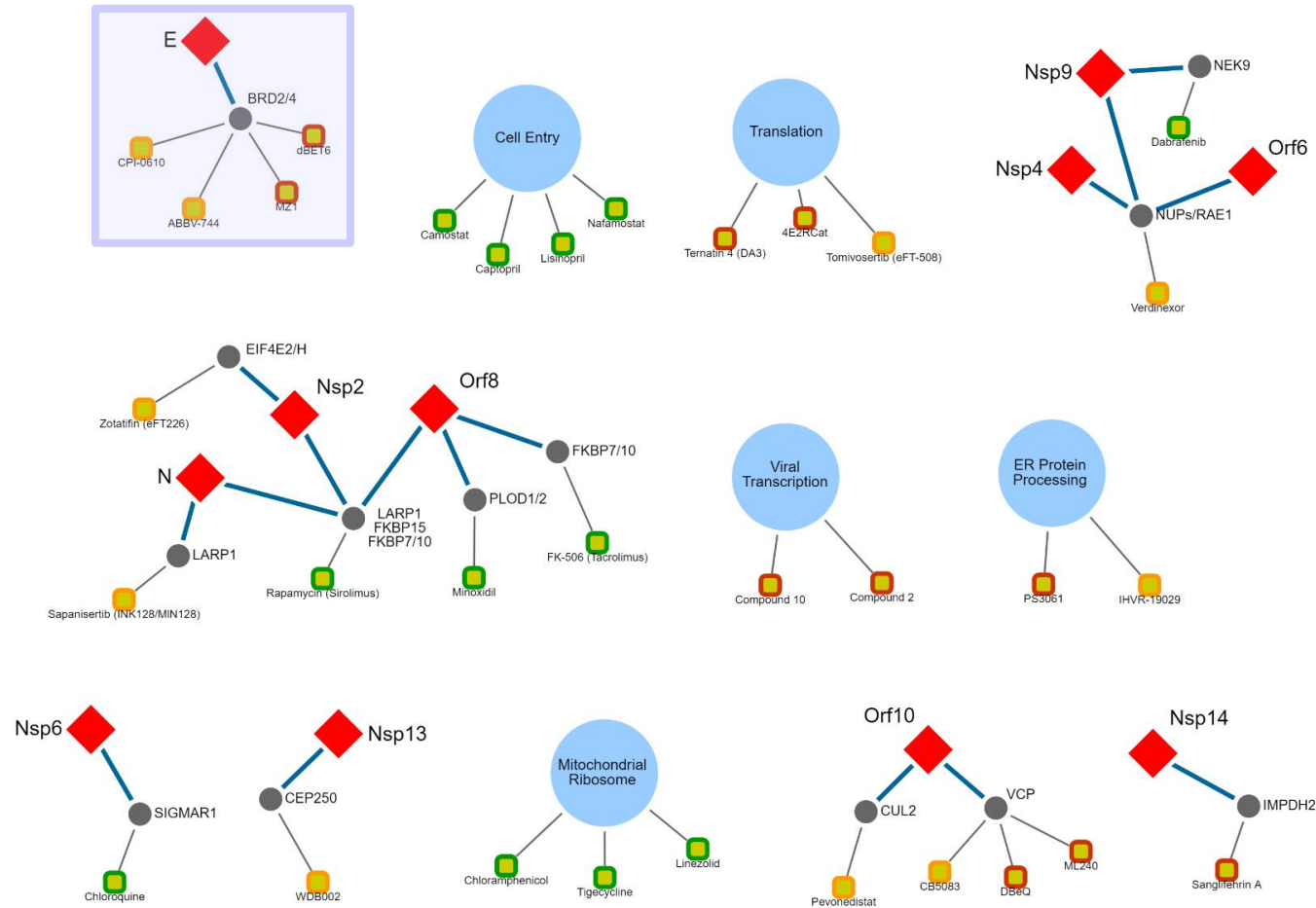
Rights Holder: Nevan Krogan
Reference: bioRxiv - doi: 10.1101/2020.03.22.002386
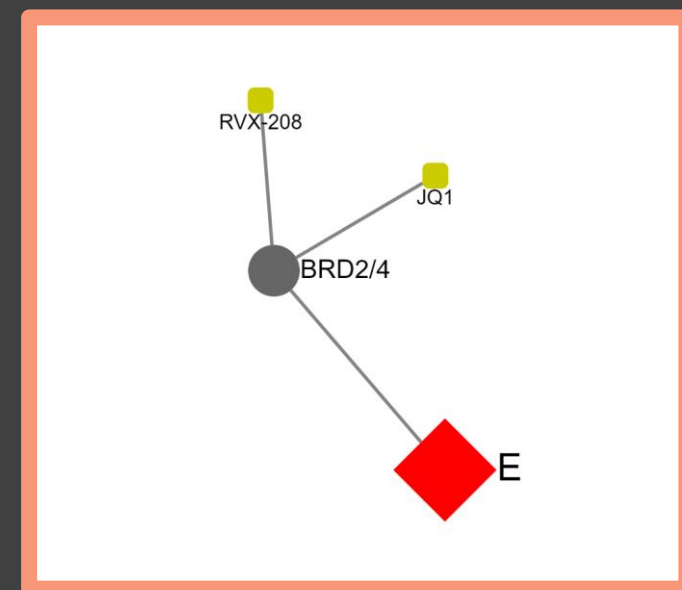
Expert-identified drugs and reagents that modulate SARS-CoV-2 interactors

Rights Holder: Nevan Krogan | Reference: bioRxiv - doi: 10.1101/2020.03.22.002386

Supplementary Table 4

| COMPOUND NAME	HUMAN GENE/PROCESS	VIRAL BAIT	DRUG STATUS	ACTIVITY DESCRIPTION (NM)
ABBV-744	BRD2/4	E	Clinical Trial	BRD inhibitor, K_D = 2.1
dBET6	BRD2/4	E	Pre-clinical	Degrades BRD proteins, IC_{50} < 10000
MZ1	BRD2/4	E	Pre-clinical	Degrades BRD proteins, K_D = 120-228
CPI-0610	BRD2/4	E	Clinical Trial	BRD2/4 inhibitor, BRD2 IC_{50} = 25, BRD4 IC_{50} = 18

Development Status
Green border = Approved
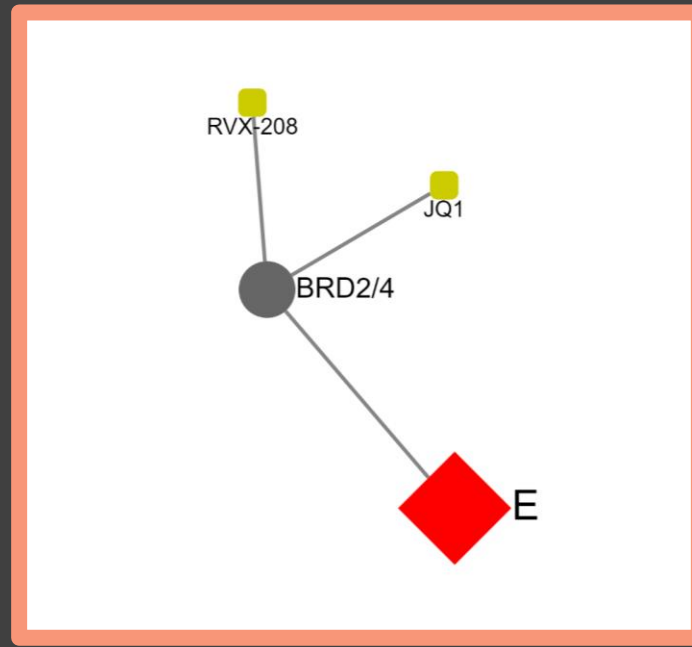**Orange border** = Clinical trial
Red border = Pre-clinical

Thicker Dark Blue edges represent host-pathogen protein-protein interactions
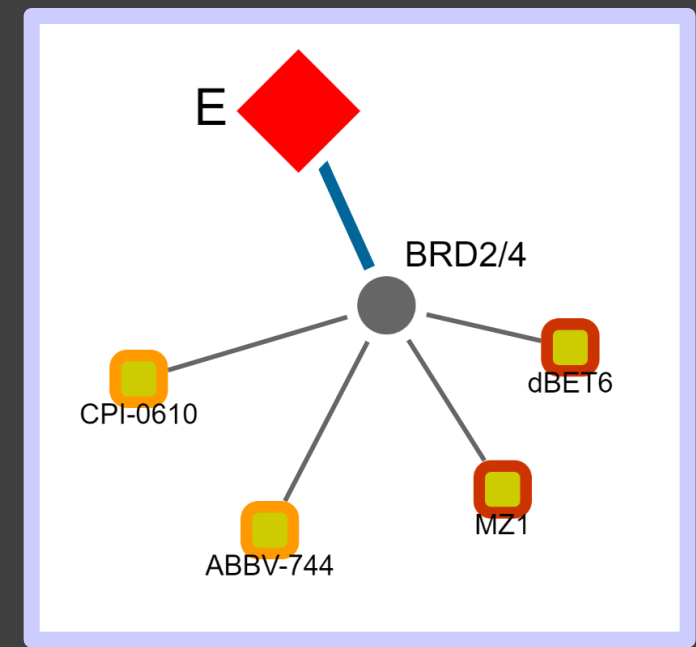
# Create a **composite SARS-CoV-2 network** showing compounds that can inhibit viral/human protein interactions
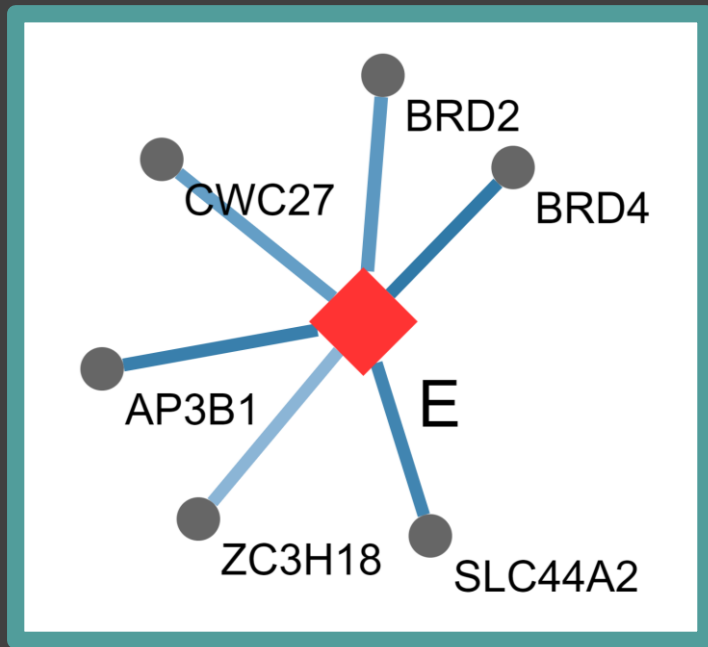
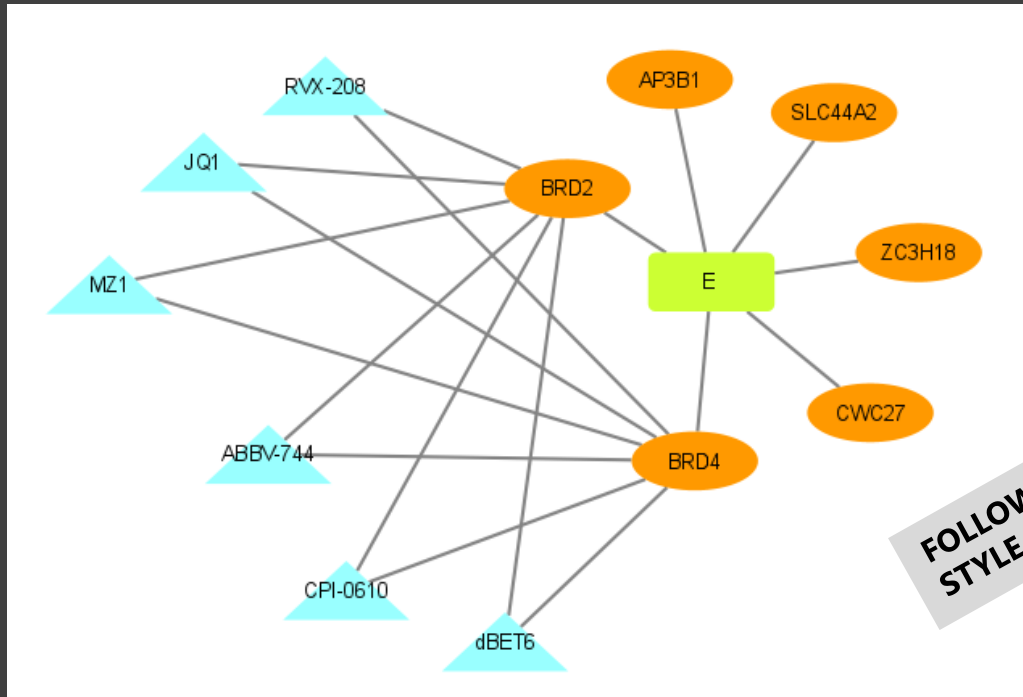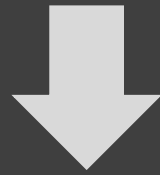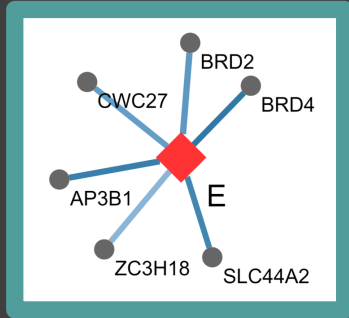**SARS-CoV-2 / Human-Protein**

Literature-derived drugs

Expert-identified drugs

Modified Data for "E" Protein - Cytoscape Model

| ACTIVITY/ MIST SCORE	HUMAN PROTEIN	UNP ID	DRUG/VIRAL PROTEIN	STATUS/VIRAL TYPE	ACTIVITY/VIRUS
2.1	BRD2	P25440	ABBV-744	Clinical Trial	BRD inhibitor, KD = 2.1
100	BRD2	P25440	dBET6	Pre-clinical	Degrades BRD proteins, IC50 < 10000
120	BRD2	P25440	MZ1	Pre-clinical	Degrades BRD proteins, KD = 120-228
25	BRD2	P25440	CPI-0610	Clinical Trial	BRD2 inhibitor, BRD2 IC50 = 25
2.1	BRD4	O60885	ABBV-744	Clinical Trial	BRD inhibitor, KD = 2.1
100	BRD4	O60885	dBET6	Pre-clinical	Degrades BRD proteins, IC50 < 10000
120	BRD4	O60885	MZ1	Pre-clinical	Degrades BRD proteins, KD = 120-228
18	BRD4	O60885	CPI-0610	Clinical Trial	BRD4 inhibitor, BRD4 IC50 = 18
40	BRD2	P25440	JQ1	Pre-clinical	BRD inhibitor, IC50 = 40-120
50	BRD2	P25440	RVX-208	Clinical Trial	BRD inhibitor, IC50 = 50-1800
40	BRD4	O60885	JQ1	Pre-clinical	BRD inhibitor, IC50 = 40-120
50	BRD4	O60885	RVX-208	Clinical Trial	BRD inhibitor, IC50 = 50-1800
0.906592876	BRD2	P25440	E	SP	SARS-CoV-2
0.97848835	BRD4	O60885	E	SP	SARS-CoV-2
0.89310916	CWC27	Q6UX04	E	SP	SARS-CoV-2
0.950342834	SLC44A2	Q8IWA5	E	SP	SARS-CoV-2
0.796415039	ZC3H18	Q86VM9	E	SP	SARS-CoV-2
0.963550095	AP3B1	O00203	E	SP	SARS-CoV-2
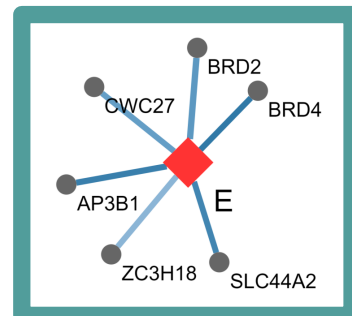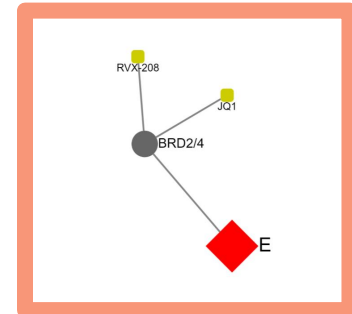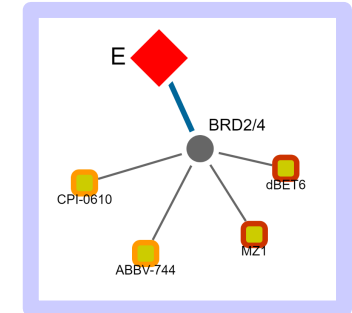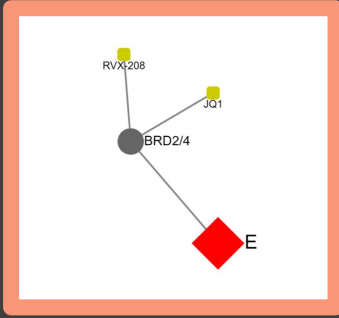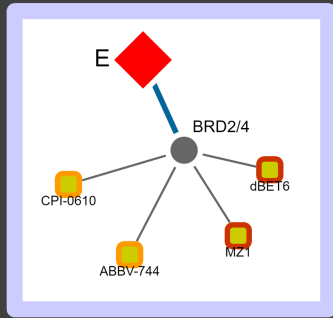
FOLLOWING STYLE PROTOCOLS
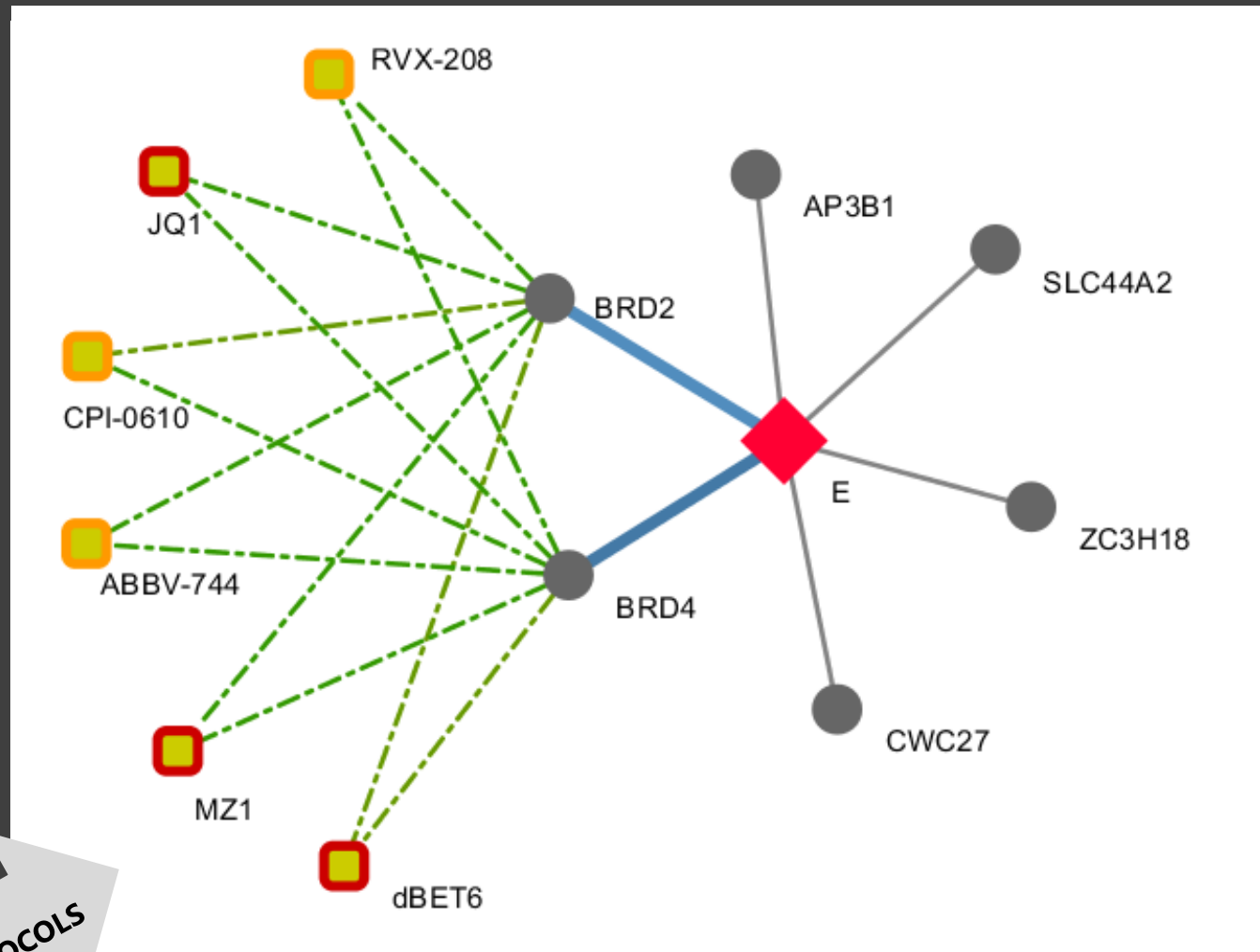
"E" Protein Network with Compound Nodes
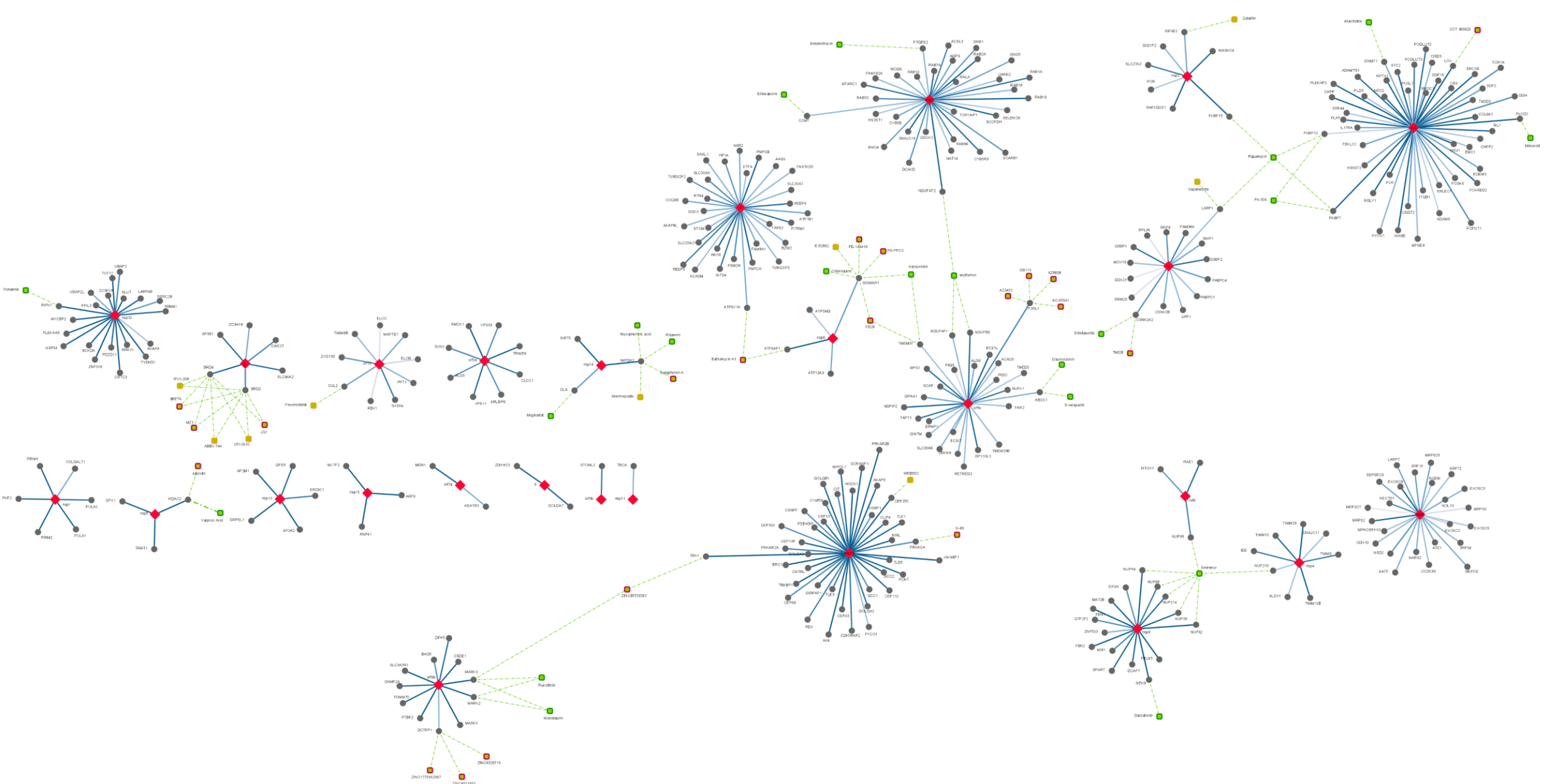
SARS-CoV-2 Full Recreated Network with Compound Nodes
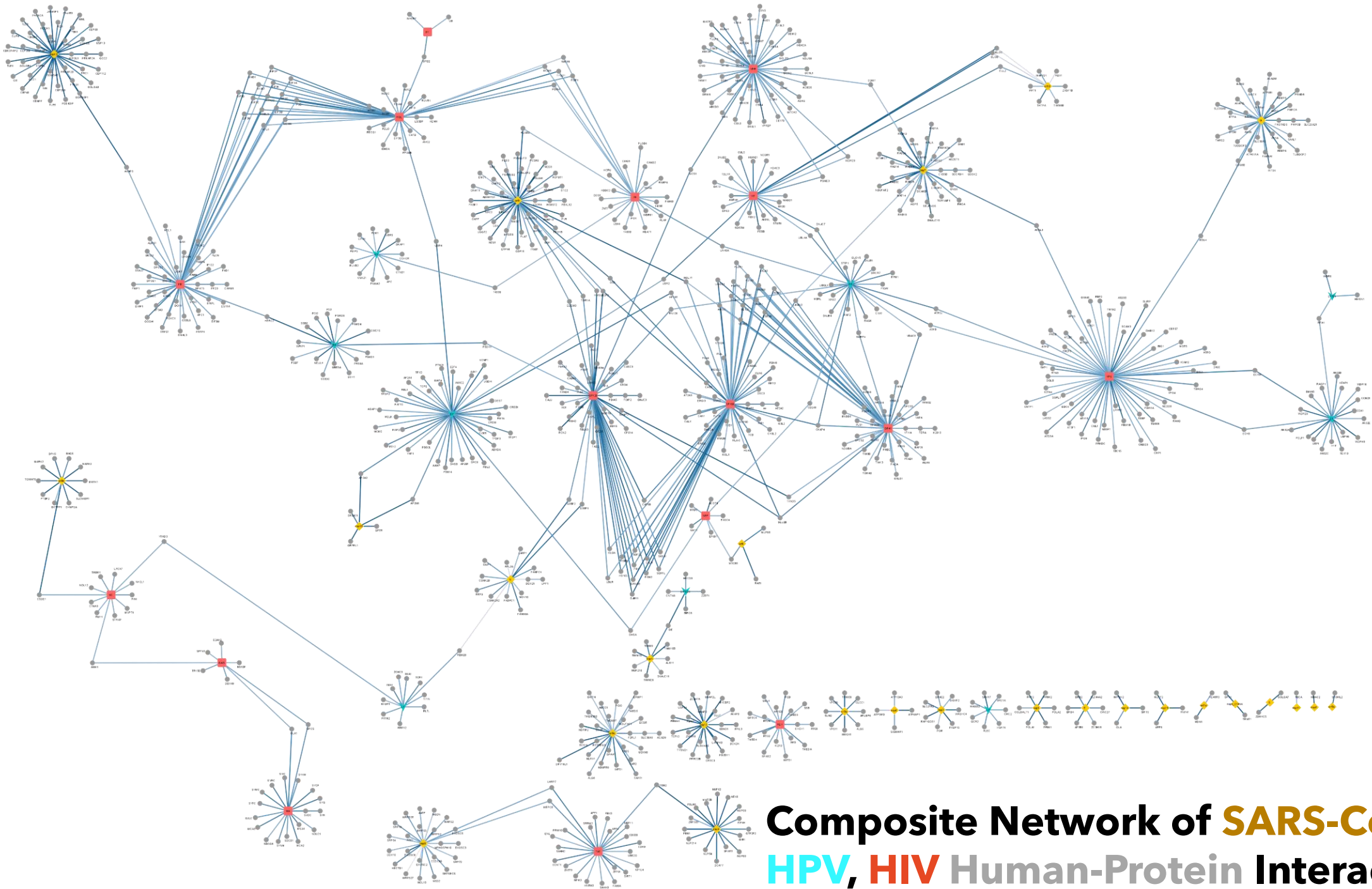
Composite Network of SARS-CoV-2, HPV, HIV Human-Protein Interactions

- Build a **database** of viral/human proteins and compounds
- Run **SQL queries**

Basic SQL structure:

SELECT *column_1, column_2, etc.*
FROM *table*
WHERE *condition*;

List human proteins interacting with "E":

SELECT Main.[Viral Protein], Main.[Human Gene]
FROM Main
WHERE ((Main.[Viral Protein])="E") And
((Main.[MIST]>0));
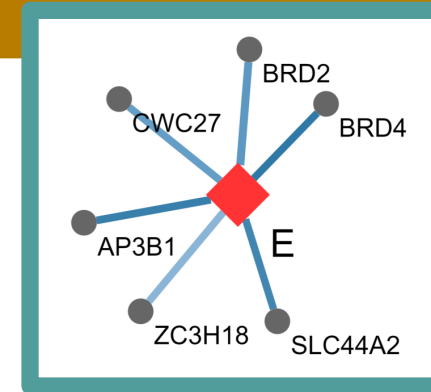
Viral Protein	Human Gene
E	AP3B1
E	BRD2
E	BRD4
E	CWC27
E	SLC44A2
E	ZC3H18

SQL Query

"Switch" operator assigns discrete values for range of MIST Scores in new column called "MISTRange"
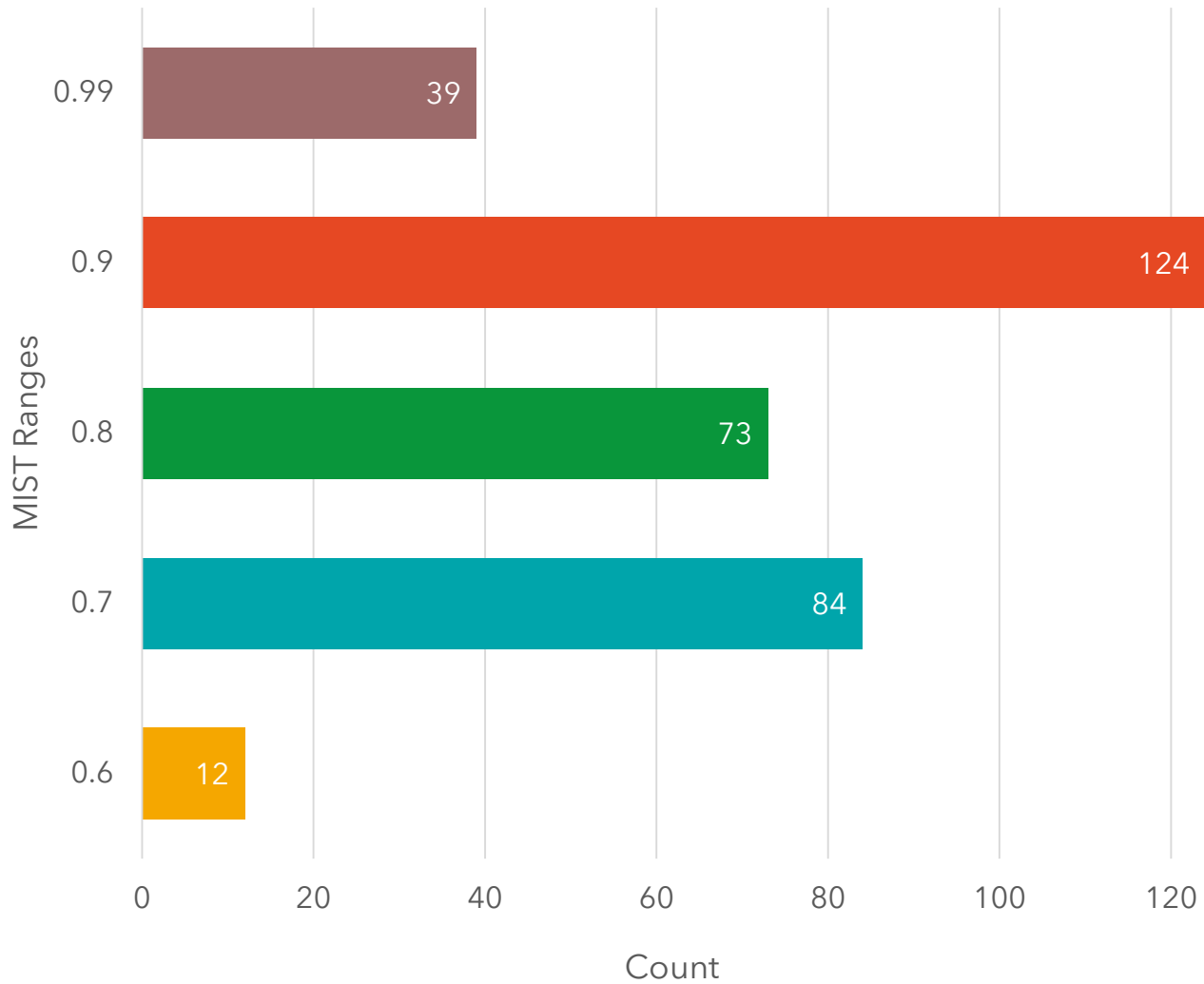
- Anything between 0.6 and less than 0.7 is assigned "0.6"
- Anything between 0.7 and less than 0.8 is assigned to "0.7"

```
SELECT DISTINCT Main.[MIST],

Switch(
        Main.[MIST] < 0.70, 0.60,
        Main.[MIST] < 0.80, 0.70,
        Main.[MIST] < 0.90, 0.80,
        Main.[MIST] < 0.99, 0.90,
        Main.[MIST] < 1, 0.99
        ) AS MISTRange

FROM Main;
```

MIST	MISTRange
0.618948613	0.6
0.628043125	0.6
0.631167513	0.6
0.6363728	0.6
0.63921796	0.6
0.647089409	0.6
0.655233005	0.6
0.660192254	0.6
0.669574426	0.6
0.672587659	0.6
0.701764404	0.7
0.703620586	0.7
0.704291901	0.7
0.705393478	0.7
0.706328526	0.7
0.706466834	0.7
0.710174697	0.7
0.710961769	0.7
0.71407894	0.7
0.717059992	0.7
0.717265558	0.7
0.718398295	0.7
0.720285746	0.7
0.720456009	0.7
0.721790867	0.7

Range of MIST Scores


```
SELECT

Sum(IIf((MISTRange.MISTRange)=0.6,1,0)) AS Pt6,
Sum(IIf((MISTRange.MISTRange)=0.7,1,0)) AS Pt7,
Sum(IIf((MISTRange.MISTRange)=0.8,1,0)) AS Pt8,
Sum(IIf((MISTRange.MISTRange)=0.9,1,0)) AS Pt9,
Sum(IIf((MISTRange.MISTRange)=0.99,1,0)) AS Pt99

FROM MISTRange

WHERE (((MISTRange.MISTRange) In (0.7,0.8,0.9,0.6,0.99)));
```

Pt6	Pt7	Pt8	Pt9	Pt99
12	84	73	124	39

SQL: Common Human Genes In SARS-CoV-2 and HPV Where MIST scores >= 0.6

	A	B	C	D	E
1	Human Gene	SARS-CoV-2 Viral Protein	SARS-CoV-2 MIST (HEK293)	HPV Viral Protein	HPV MIST (C33A)
2	HDAC2	nsp5	0.993708403	L2	0.685510627
3	AP2A2	nsp10	0.99112813	E7	0.882203347
4	AP2M1	nsp10	0.982905884	E7	0.853236047
5	WFS1	orf9c	0.955124813	E5	0.685510241
6	IDE	nsp4	0.918031442	E8^E2C	0.974307196
7	RTN4	M	0.873826097	E5	0.685510371
8	AAR2	M	0.801486724	E6	0.685511079
9	ACADM	M	0.724348569	E7	0.685511333
10	EMC1	orf8	0.723777507	E5	0.685510216
11	AKAP8	nsp12	0.717059992	L1	0.685510281
12	RBM28	N	0.628043125	L1	0.816328465

SELECT DISTINCT

Main.[Human Gene], Main.[Viral Protein], Main.[MIST], HPV.[Viral Protein], HPV.[MIST]

FROM Main

INNER JOIN HPV ON Main.[Human Gene] = HPV.[Human Gene]

WHERE Main.MIST>=0.6 And HPV.MIST>=0.6;

SQL: Common Human Genes In SARS-CoV-2 and HIV Where MIST scores >= 0.6

	A	B	C	D	E	F
1	Human Gene	SARS-CoV-2 Viral Protein	SARS-CoV-2 MIST HEK293	HIV Viral Protein	HIV MIST HEK293	HIV MIST JurKat
2	NGLY1	orf8	0.993220351	GP160	0	0.834
3	NGLY1	orf8	0.993220351	GP41	0	0.756
4	PLOD2	orf8	0.992483418	IN	0.754	0
5	FBN2	nsp9	0.991012329	TAT	0.838	0
6	AKAP9	nsp13	0.990813809	PR	0	0.787
7	CSDE1	orf9b	0.988959751	NC	0.812	0
8	G3BP2	N	0.958133672	GP120	0	0.838
9	G3BP1	N	0.95331626	GP120	0	0.784
10	OS9	orf8	0.931530938	GP160	0	0.915
11	ACSL3	nsp7	0.897068932	VPU	0.821	0.058
12	TOR1A	orf8	0.879789245	GP120	0	0.846
13	RTN4	M	0.873826097	VPU	0.809	0.819
14	MTCH1	orf6	0.851173737	NEF	0.752	0
15	SDF2	orf8	0.826107348	GP120	0.856	0.981
16	SDF2	orf8	0.826107348	GP160	0.829	0.931
17	EDEM3	orf8	0.81834398	GP120	0	0.898
18	CUL2	orf10	0.818290141	VIF	0.776	0.897
19	LARP7	nsp8	0.812479682	TAT	0.54	0.915
20	MEPCE	nsp8	0.790978117	TAT	0.692	0.857
21	HYOU1	orf8	0.77235306	GP120	0.775	0.96
22	COMT	nsp7	0.745231765	VPR	0.775	0.244
23	LMAN2	nsp7	0.725773983	GP160	0	0.839
24	LMAN2	nsp7	0.725773983	GP41	0.86	0
25	ELOB	orf10	0.655233005	VIF	0.96	0.928
26	ELOC	orf10	0.618948613	VIF	0.956	0.972

SELECT DISTINCT

Main.[Human Gene], Main.[Viral Protein], Main.[MIST],
      HIV.[Viral Protein], HIV.[MIST HEK293], HIV.[MIST JurKat]

FROM Main

INNER JOIN HIV ON Main.[Human Gene] = HIV.[Human Gene]
WHERE Main.MIST>=0.6 And
      (HIV.[MIST HEK293]>=0.6 OR HIV.[MIST JurKat]>=0.6);

Find the common human proteins that interact with SARS-CoV-2, HIV, and HPV?

```
SELECT DISTINCT

Main.[Human Gene],
    Main.[Viral Protein], Main.[MIST],
    HIV.[Viral Protein], HIV.[MIST HEK293], HIV.[MIST JurKat]
    HPV.[Viral Protein], HPV.[MIST],

FROM (Main
    INNER JOIN HPV ON Main.[Human Gene] = HPV.[Human Gene])

INNER JOIN HIV ON HPV.[Human Gene] = HIV.[Human Gene];
```
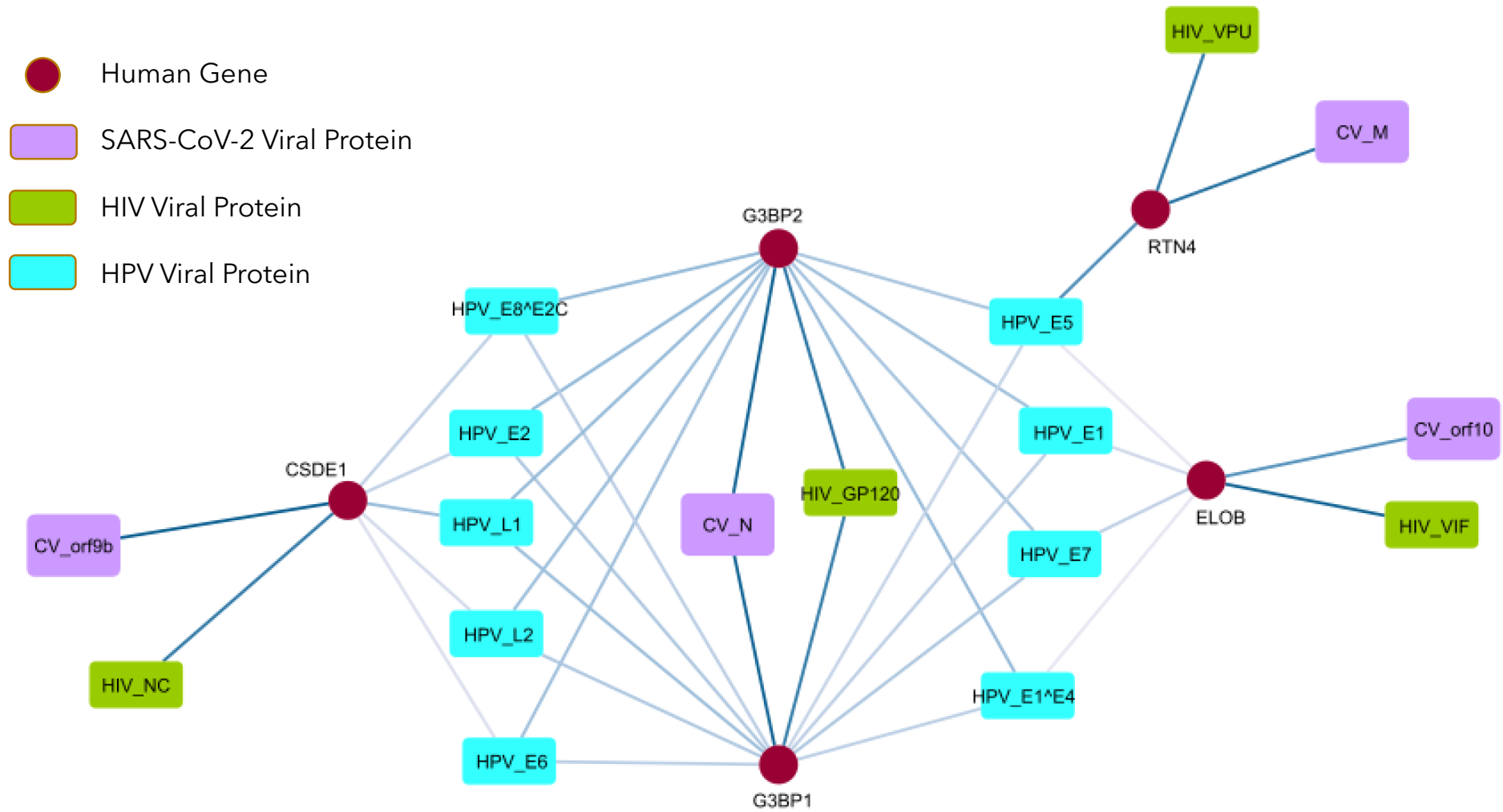
Common Gene Interactions In SARS-CoV-2, HIV, HPV

	Human Gene	SARS-CoV-2 Viral Protein	SARS-CoV-2 MIST HEK293	HIV Viral Protein	HIV MIST HEK293	HIV MIST JurKat	HPV Viral Protein	HPV MIST (C33A)
2	**CSDE1**	orf9b	0.988959751	NC	0.812	0	E2	0.176316087
3	CSDE1	orf9b	0.988959751	NC	0.812	0	E6	0.0698267
4	CSDE1	orf9b	0.988959751	NC	0.812	0	E8^E2C	0.186285804
5	CSDE1	orf9b	0.988959751	NC	0.812	0	L1	0.356805069
6	CSDE1	orf9b	0.988959751	NC	0.812	0	L2	0.103342845
7	**G3BP2**	N	0.958133672	GP120	0	0.838	E1	0.322921149
8	G3BP2	N	0.958133672	GP120	0	0.838	E1^E4	0.363596661
9	G3BP2	N	0.958133672	GP120	0	0.838	E2	0.383268897
10	G3BP2	N	0.958133672	GP120	0	0.838	E5	0.290177749
11	G3BP2	N	0.958133672	GP120	0	0.838	E6	0.317960495
12	G3BP2	N	0.958133672	GP120	0	0.838	E7	0.289612609
13	G3BP2	N	0.958133672	GP120	0	0.838	E8^E2C	0.385217699
14	G3BP2	N	0.958133672	GP120	0	0.838	L1	0.365959601
15	G3BP2	N	0.958133672	GP120	0	0.838	L2	0.32994147
16	**G3BP1**	N	0.95331626	GP120	0	0.784	E1	0.196739252
17	G3BP1	N	0.95331626	GP120	0	0.784	E1^E4	0.200569845
18	G3BP1	N	0.95331626	GP120	0	0.784	E2	0.263651559
19	G3BP1	N	0.95331626	GP120	0	0.784	E5	0.161791186
20	G3BP1	N	0.95331626	GP120	0	0.784	E6	0.20952895
21	G3BP1	N	0.95331626	GP120	0	0.784	E7	0.26312747
22	G3BP1	N	0.95331626	GP120	0	0.784	E8^E2C	0.204416776
23	G3BP1	N	0.95331626	GP120	0	0.784	L1	0.335680758
24	G3BP1	N	0.95331626	GP120	0	0.784	L2	0.258371514
25	**RTN4**	M	0.873826097	VPU	0.809	0.819	E5	0.685510371
26	**ELOB**	orf10	0.655233005	VIF	0.96	0.928	E1	0.120568043
27	ELOB	orf10	0.655233005	VIF	0.96	0.928	E1^E4	0.016731554
28	ELOB	orf10	0.655233005	VIF	0.96	0.928	E5	0.016963384
29	ELOB	orf10	0.655233005	VIF	0.96	0.928	E7	0.189816337

Cytoscape: Common Gene Interactions In SARS-CoV-2, HIV, HPV

Common Gene Interactions In SARS-CoV-2, HIV, HPV Where MIST scores >= 0.6

	A	B	C	D	E	F	G	H	I
1	UNP ID	Human Gene	SARS-CoV-2 Viral Protein	SARS-CoV-2 MIST HEK293	HIV Viral Protein	HIV MIST HEK293	HIV MIST JurKat	HPV Viral Protein	HPV MIST (C33A)
2	Q9NQC3	RTN4	M	0.873826097	VPU	0.809	0.819	E5	0.685510371

SELECT DISTINCT Main.[Human Gene], Main.[UNP ID],

    Main.[Viral Protein], Main.[MIST],
    HIV.[Viral Protein], HIV.[MIST HEK293], HIV.[MIST JurKat],
    HPV.[Viral Protein], HPV.[MIST]

FROM (Main INNER JOIN HIV ON Main.[Human Gene] = HIV.[Human Gene])
INNER JOIN HPV ON HIV.[Human Gene] = HPV.[Human Gene]

WHERE Main.[MIST]>=0.6 AND HPV.[MIST]>=0.6
    AND (HIV.[MIST HEK293]>=0.6 OR HIV.[MIST JurKat]>=0.6);

Acknowledgements