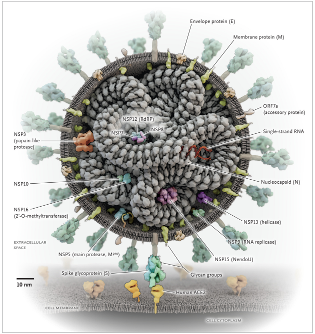
CS 3824: Identifying Human Interactors of SARS-CoV-2 Proteins Using Interpretable Network Propagation

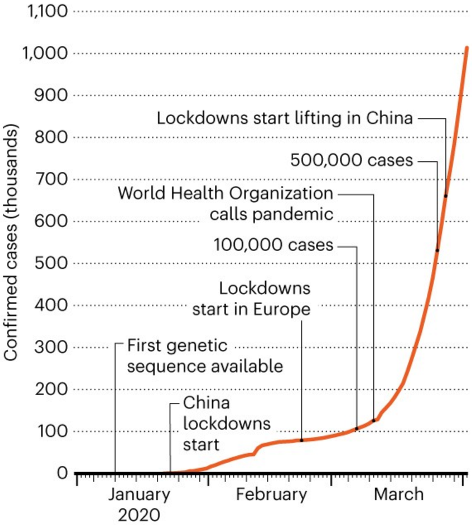
T. M. Murali

October 4, 6, 11, 2022



ONE MILLION INFECTIONS

There have now been more than one million confirmed cases of the coronavirus worldwide.



Data correct as of 3 April 2020

MATTER

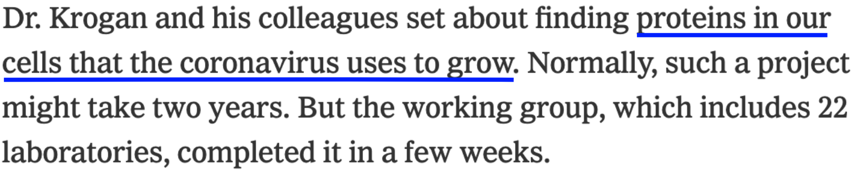
Hundreds of Scientists Scramble to Find a Coronavirus Treatment

In an ambitious international collaboration, researchers have "mapped" proteins in the coronavirus and identified 50 drugs to test against it.

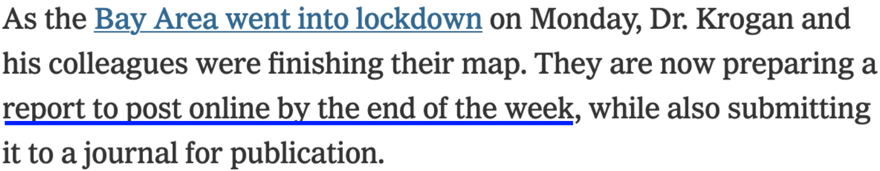
By Carl Zimmer

March 17, 2020

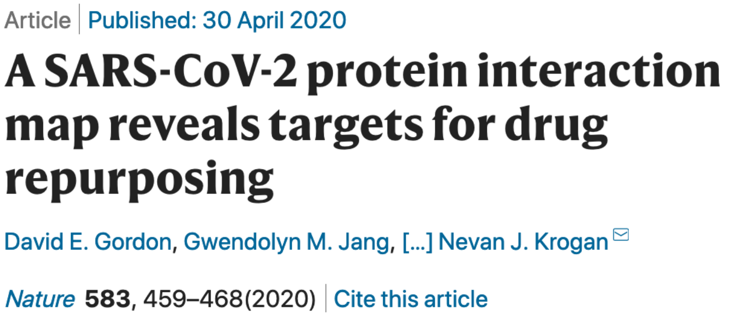
271



Dr. Krogan and his colleagues set about finding proteins in our cells that the coronavirus uses to grow. Normally, such a project might take two years. But the working group, which includes 22 laboratories, completed it in a few weeks.



As the Bay Area went into lockdown on Monday, Dr. Krogan and his colleagues were finishing their map. They are now preparing a report to post online by the end of the week, while also submitting it to a journal for publication.



Article | Published: 30 April 2020

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

David E. Gordon, Gwendolyn M. Jang, [...] Nevan J. Krogan

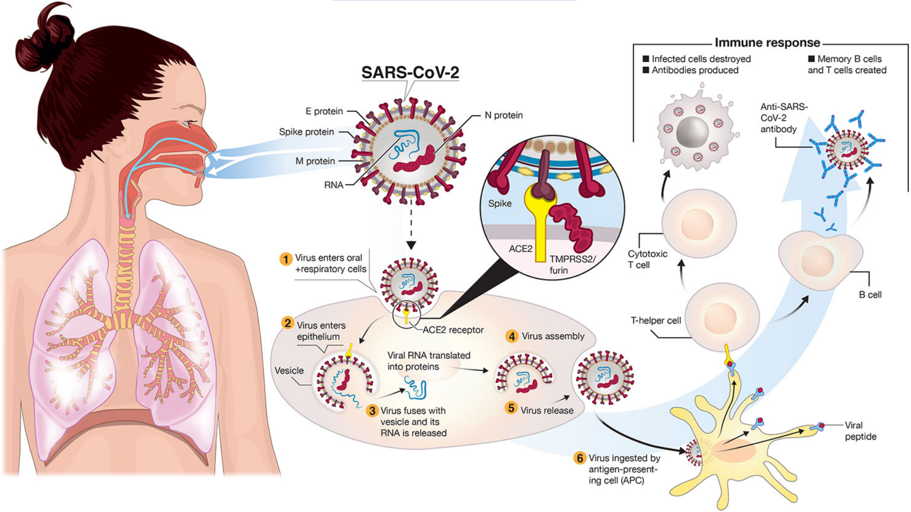
Nature **583**, 459–468(2020) | Cite this article

- Supplementary file contained list of 332 human proteins that interact with SARS-CoV-2.

Gordon et al. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals [...]. Nature, April 30, 2020.

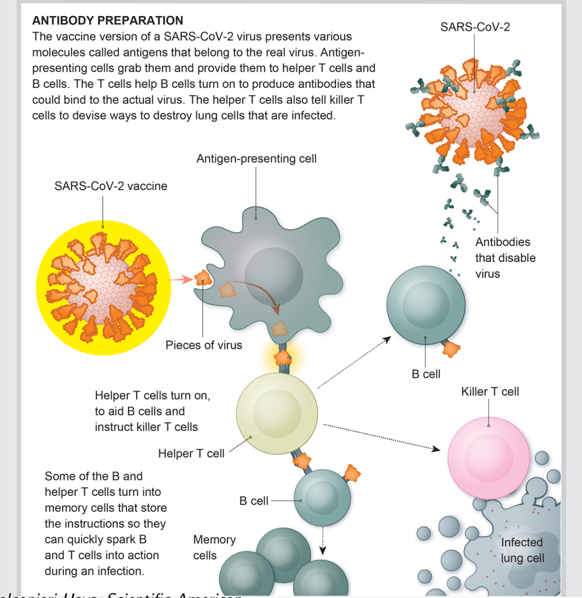
SARS-CoV-2 Life Cycle

▸ Video on SARS-CoV-2 life cycle



Funk, Laferrière, Ardakani, "A Snapshot of the Global Race for Vaccines Targeting SARS-CoV-2 and the COVID-19 Pandemic," Frontiers in Pharmacology, 11, 937, 2020

Vaccines



ANTIBODY PREPARATION

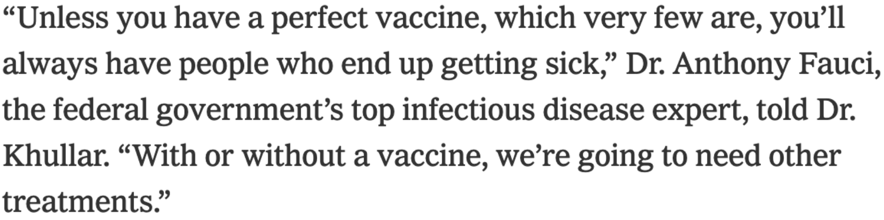
The vaccine version of a SARS-CoV-2 virus presents various molecules called antigens that belong to the real virus. Antigen-presenting cells grab them and provide them to helper T cells and B cells. The T cells help B cells turn on to produce antibodies that could bind to the actual virus. The helper T cells also tell killer T cells to devise ways to destroy lung cells that are infected.

Helper T cells turn on, to aid B cells and instruct killer T cells

Some of the B and helper T cells turn into memory cells that store the instructions so they can quickly spark B and T cells into action during an infection.

Image credit: Veronica Falconieri Hays, Scientific American.

Vaccines May Not Be Enough

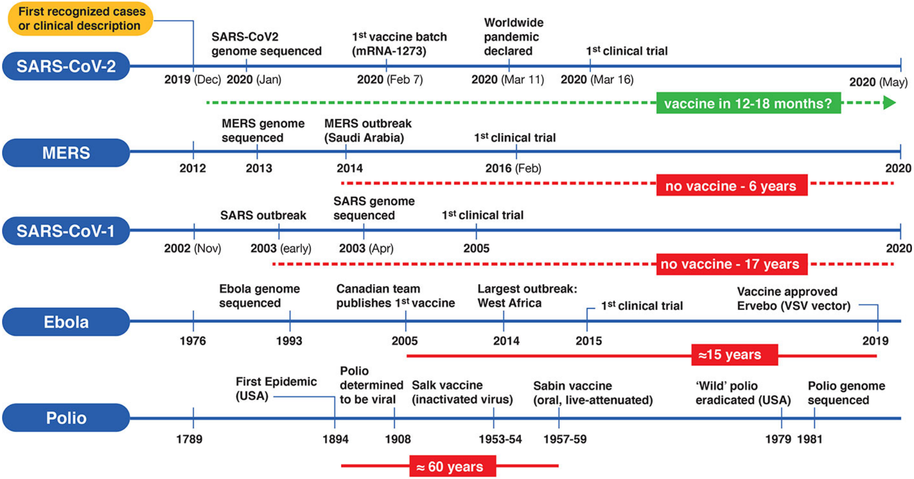


"Unless you have a perfect vaccine, which very few are, you'll always have people who end up getting sick," Dr. Anthony Fauci, the federal government's top infectious disease expert, told Dr. Khullar. "With or without a vaccine, we're going to need other treatments."

- Vaccines are not universal, can have side effects.

Khullar, "It Will Take More Than a Vaccine to Beat COVID-19," The New Yorker, September 8, 2020.

Vaccines May Not Be Enough



First recognized cases or clinical description

SARS-CoV-2
- 2019 (Dec)
- 2020 (Jan): SARS-CoV2 genome sequenced
- 2020 (Feb 7): 1st vaccine batch (mRNA-1273)
- 2020 (Mar 11): Worldwide pandemic declared
- 2020 (Mar 16): 1st clinical trial
- 2020 (May)
- vaccine in 12-18 months?

MERS
- 2012
- 2013: MERS genome sequenced
- 2014: MERS outbreak (Saudi Arabia)
- 2016 (Feb): 1st clinical trial
- 2020
- no vaccine - 6 years

SARS-CoV-1
- 2002 (Nov)
- 2003 (early): SARS outbreak
- 2003 (Apr): SARS genome sequenced
- 2005: 1st clinical trial
- 2020
- no vaccine - 17 years

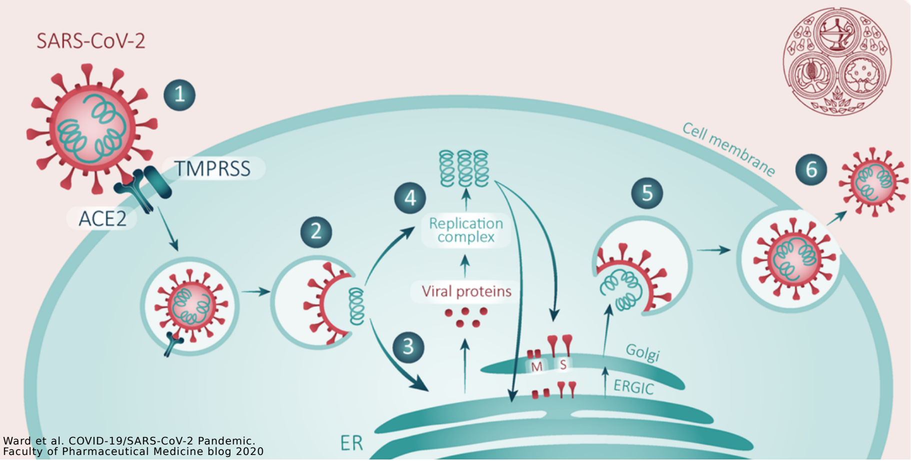
Ebola
- 1976
- 1993: Ebola genome sequenced
- 2005: Canadian team publishes 1st vaccine
- 2014: Largest outbreak: West Africa
- 2015: 1st clinical trial
- 2019: Vaccine approved Ervebo (VSV vector)
- ≈15 years

Polio
- 1789
- 1894: First Epidemic (USA)
- 1908: Polio determined to be viral
- 1953-54: Salk vaccine (inactivated virus)
- 1957-59: Sabin vaccine (oral, live-attenuated)
- 1979: 'Wild' polio eradicated (USA)
- 1981: Polio genome sequenced
- ≈ 60 years

- Vaccines are not universal, can have side effects.
- Worldwide production and distribution of vaccines takes many years.

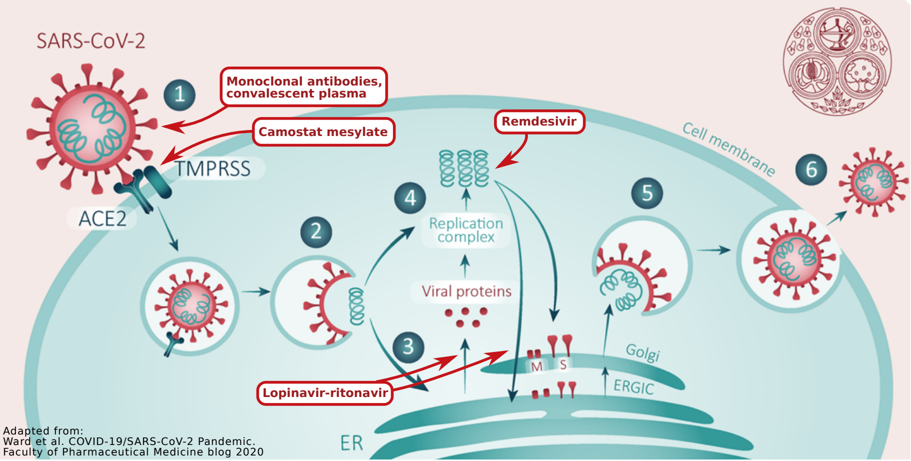
Funk, Laferrière, Ardakani, "A Snapshot of the Global Race for Vaccines Targeting SARS-CoV-2 and the COVID-19 Pandemic," Frontiers in Pharmacology, 11, 937, 2020

SARS-CoV-2 Life Cycle



Ward et al. COVID-19/SARS-CoV-2 Pandemic.
Faculty of Pharmaceutical Medicine blog 2020

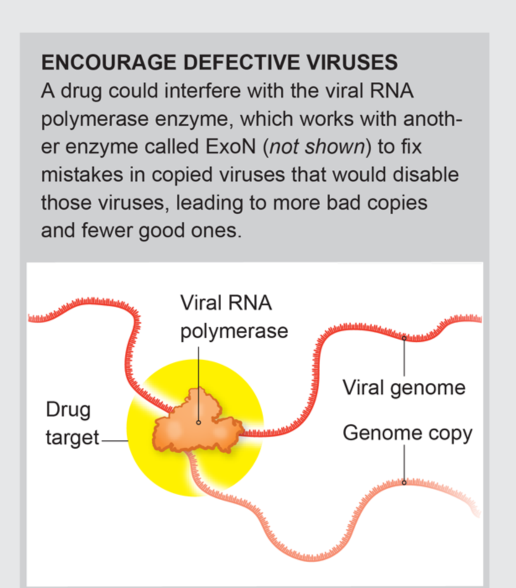
Drugs for COVID-19



Adapted from:
Ward et al. COVID-19/SARS-CoV-2 Pandemic.
Faculty of Pharmaceutical Medicine blog 2020

- Administered after the body has already been infected.

Antiviral Drugs

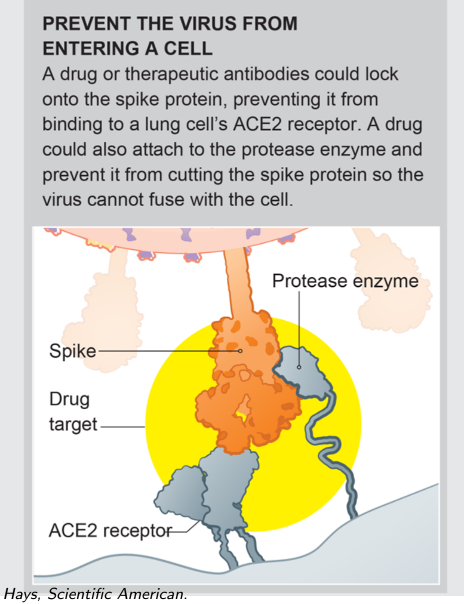


ENCOURAGE DEFECTIVE VIRUSES

A drug could interfere with the viral RNA polymerase enzyme, which works with another enzyme called ExoN (*not shown*) to fix mistakes in copied viruses that would disable those viruses, leading to more bad copies and fewer good ones.

Image credit: Veronica Falconieri Hays, Scientific American.

Antiviral Drugs

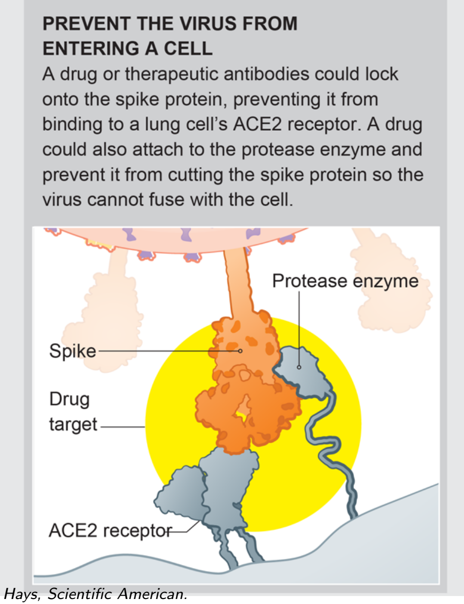


PREVENT THE VIRUS FROM ENTERING A CELL

A drug or therapeutic antibodies could lock onto the spike protein, preventing it from binding to a lung cell's ACE2 receptor. A drug could also attach to the protease enzyme and prevent it from cutting the spike protein so the virus cannot fuse with the cell.

Image credit: Veronica Falconieri Hays, Scientific American.

Host-Oriented Drugs

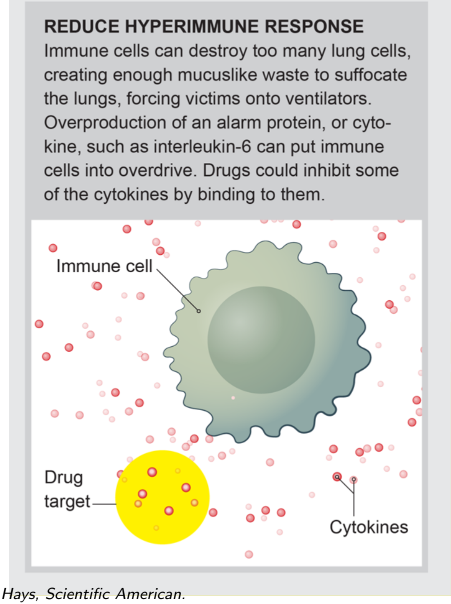


PREVENT THE VIRUS FROM ENTERING A CELL

A drug or therapeutic antibodies could lock onto the spike protein, preventing it from binding to a lung cell's ACE2 receptor. A drug could also attach to the protease enzyme and prevent it from cutting the spike protein so the virus cannot fuse with the cell.

Image credit: Veronica Falconieri Hays, Scientific American.

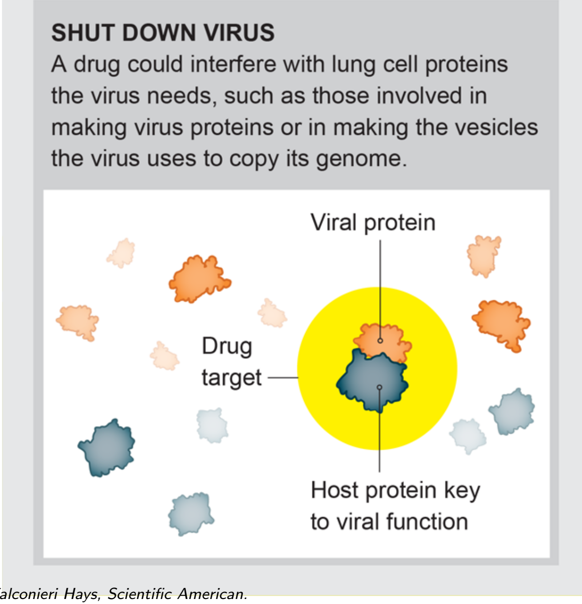
Host-Oriented Drugs



REDUCE HYPERIMMUNE RESPONSE
Immune cells can destroy too many lung cells, creating enough mucuslike waste to suffocate the lungs, forcing victims onto ventilators. Overproduction of an alarm protein, or cytokine, such as interleukin-6 can put immune cells into overdrive. Drugs could inhibit some of the cytokines by binding to them.

Image credit: Veronica Falconieri Hays, Scientific American.

Host-Oriented Drugs

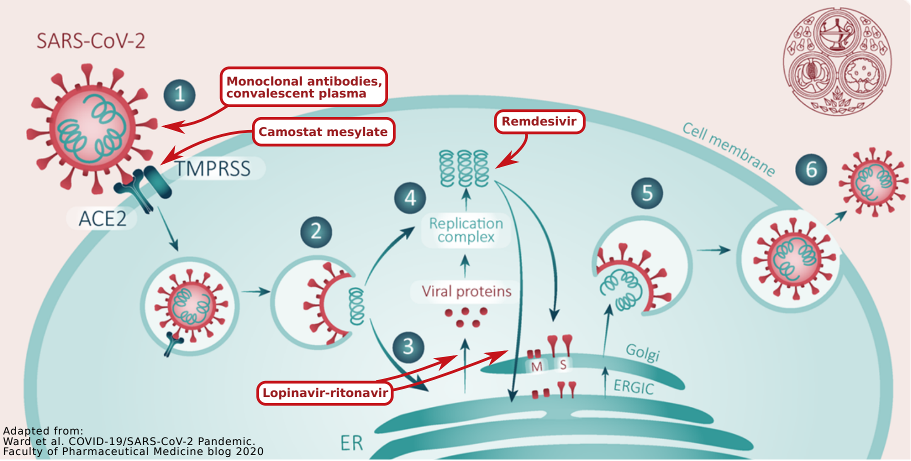


SHUT DOWN VIRUS

A drug could interfere with lung cell proteins the virus needs, such as those involved in making virus proteins or in making the vesicles the virus uses to copy its genome.

Image credit: Veronica Falconieri Hays, Scientific American.

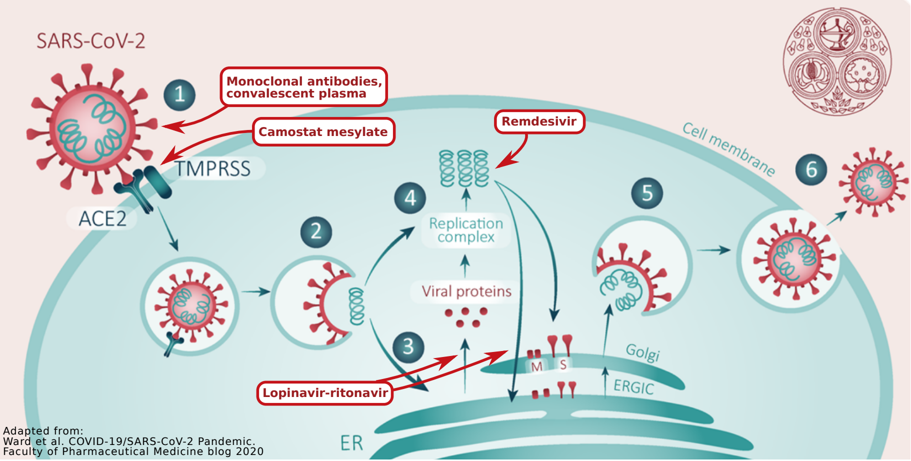
Host-Oriented Drugs for COVID-19



Adapted from:
Ward et al. COVID-19/SARS-CoV-2 Pandemic.
Faculty of Pharmaceutical Medicine blog 2020

- Viruses require host cellular processes ⇒ human proteins are viable drug targets.

Host-Oriented Drugs for COVID-19

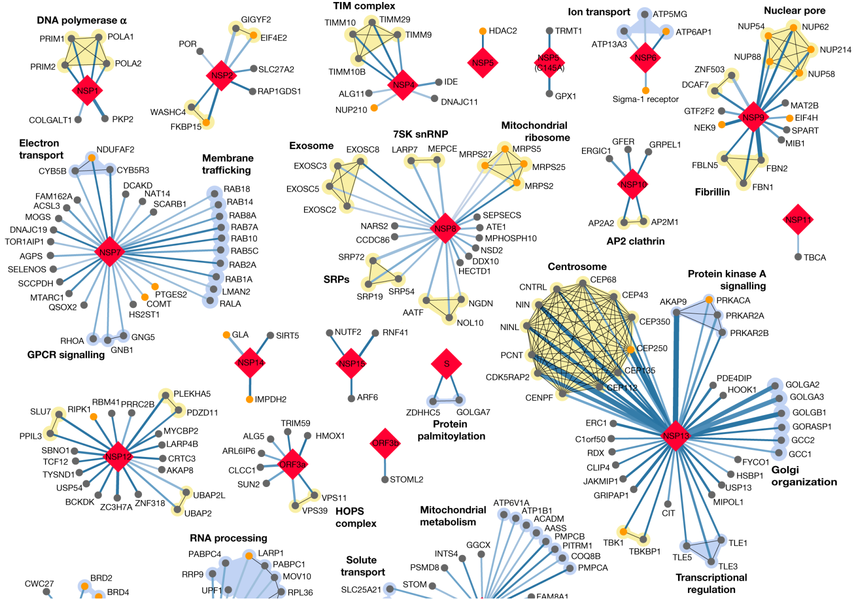


Adapted from:
Ward et al. COVID-19/SARS-CoV-2 Pandemic.
Faculty of Pharmaceutical Medicine blog 2020

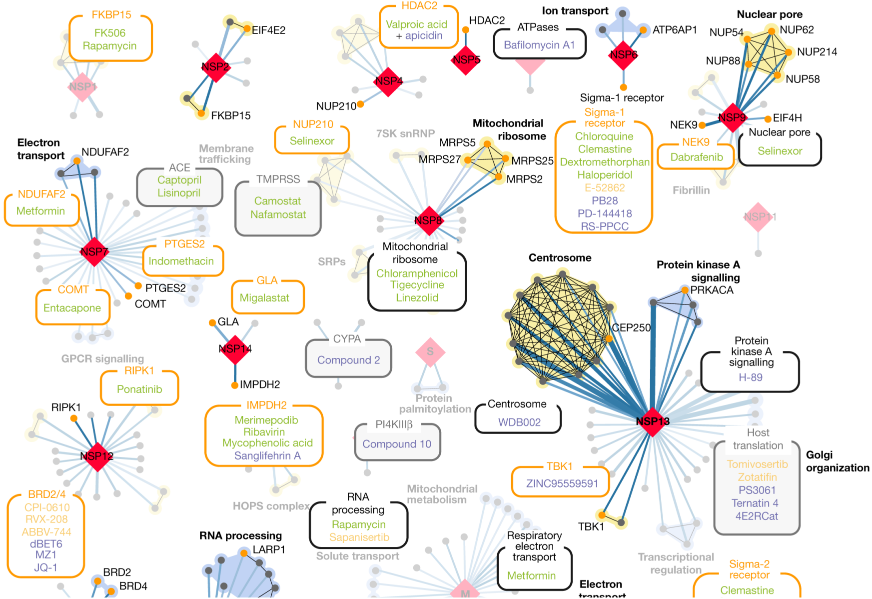
- Viruses require host cellular processes ⇒ human proteins are viable drug targets.

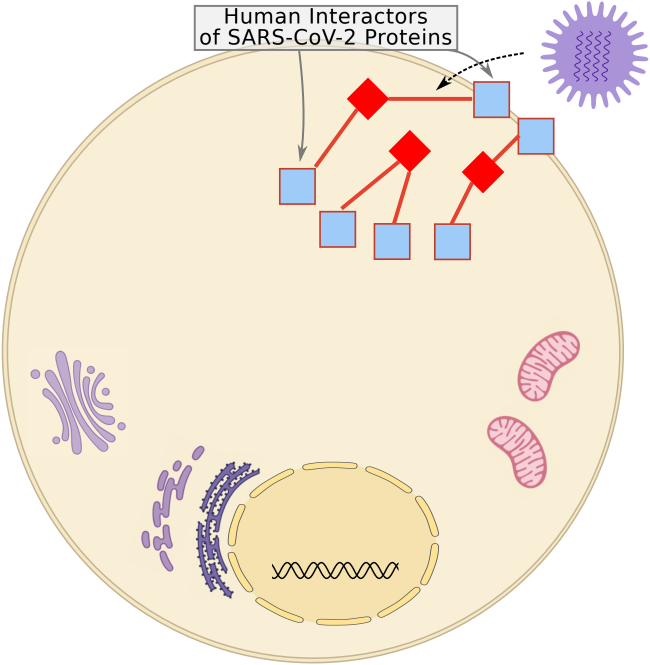
Discover human proteins and processes exploited by SARS-CoV-2.
Repurpose approved drugs for human diseases against COVID-19.

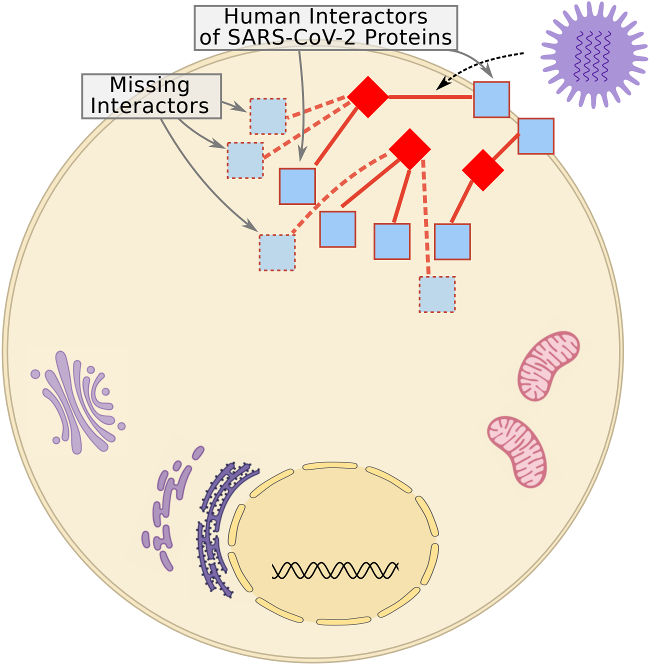
Results from Krogan's Group



Results from Krogan's Group

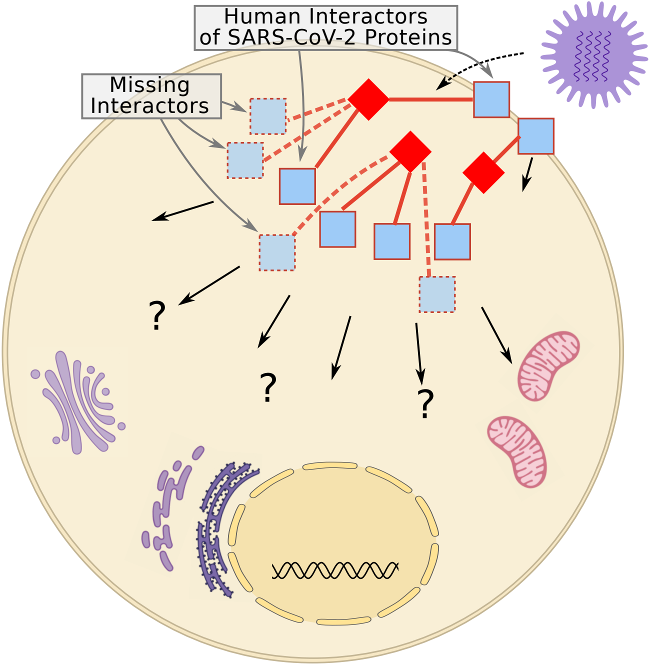


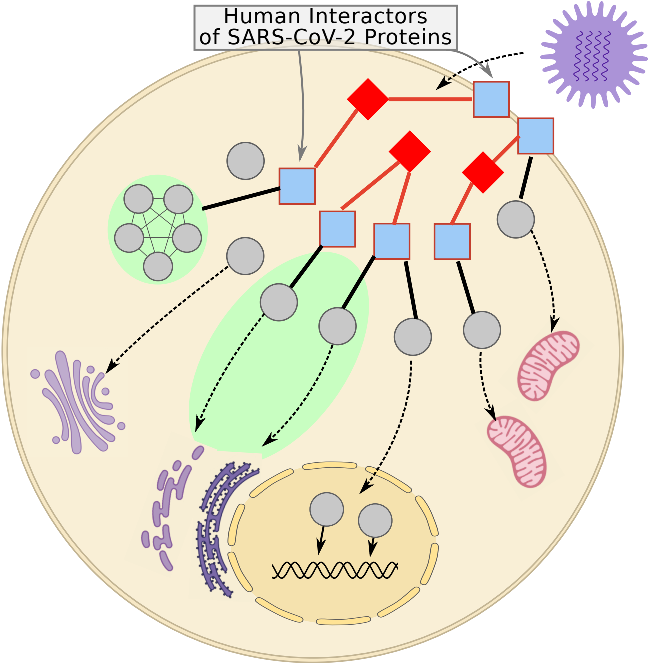

Human Interactors
of SARS-CoV-2 Proteins



Human Interactors of SARS-CoV-2 Proteins

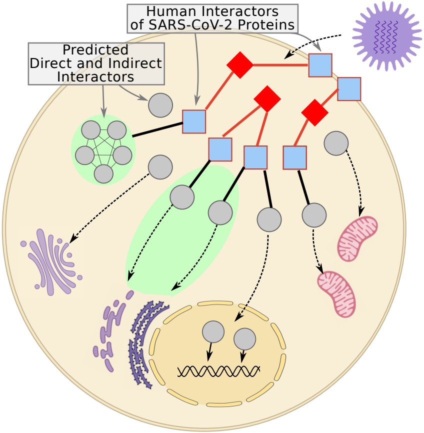
Missing Interactors


Human Interactors
of SARS-CoV-2 Proteins
Missing
Interactors
?
?
?



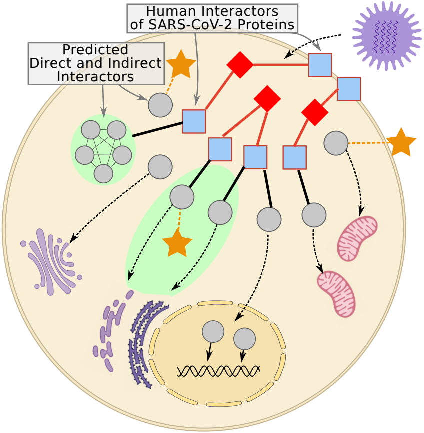
Human Interactors of SARS-CoV-2 Proteins

Goals of Our Research



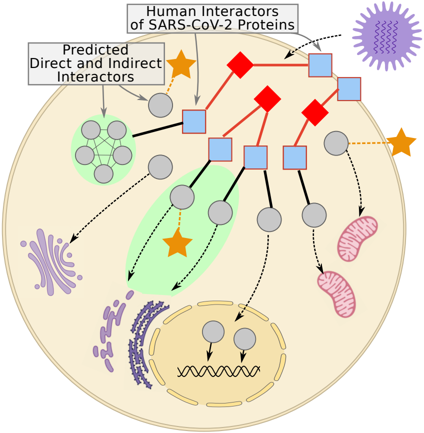
1. Prioritize additional human proteins that may directly or indirectly interact with the virus.

Goals of Our Research



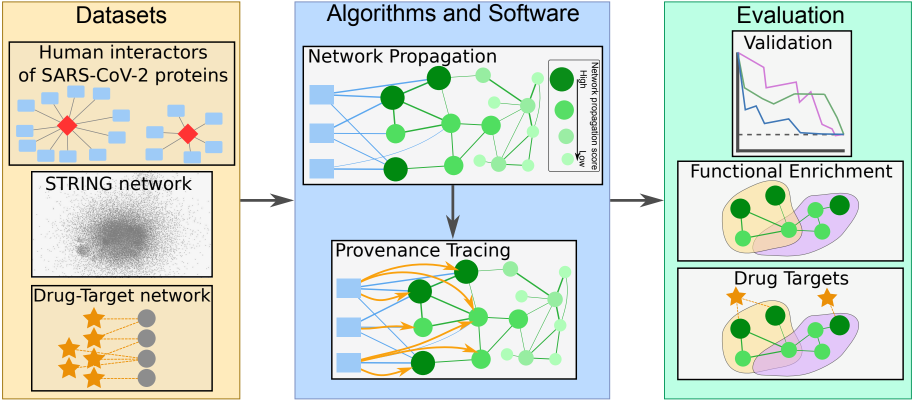
1. Prioritize additional human proteins that may directly or indirectly interact with the virus.
2. Identify drug targets among these proteins.

Goals of Our Research

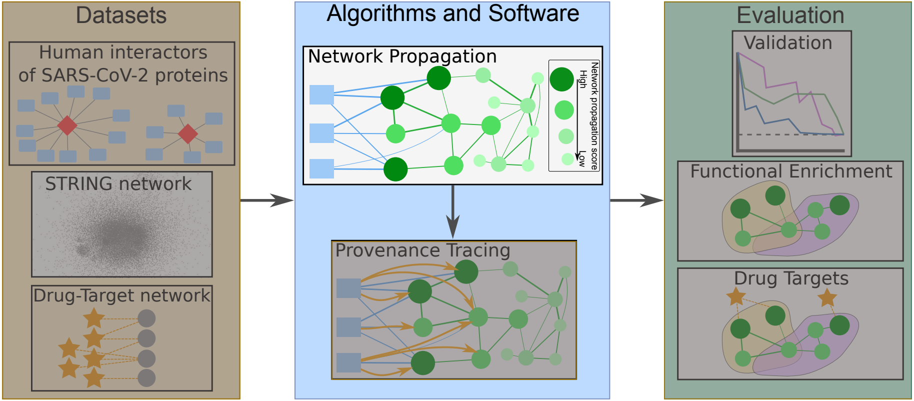


1. Prioritize additional human proteins that may directly or indirectly interact with the virus.
2. Identify drug targets among these proteins.
3. Develop strategies to explain predictions.

Approach


Datasets
Human interactors of SARS-CoV-2 proteins
STRING network
Drug-Target network
Algorithms and Software
Network Propagation
Network propagation score
High
Low
Provenance Tracing
Evaluation
Validation
Functional Enrichment
Drug Targets

Approach



Datasets

Human interactors of SARS-CoV-2 proteins

STRING network

Drug-Target network

Algorithms and Software

Network Propagation

Provenance Tracing

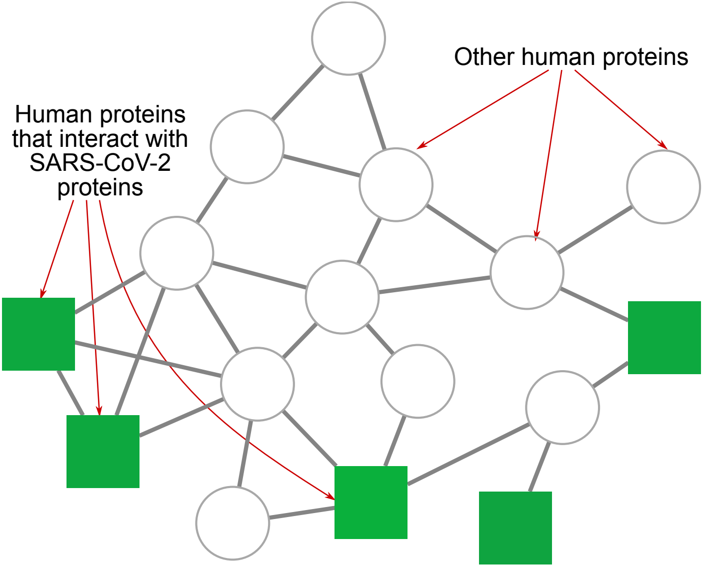
Evaluation

Validation

Functional Enrichment

Drug Targets

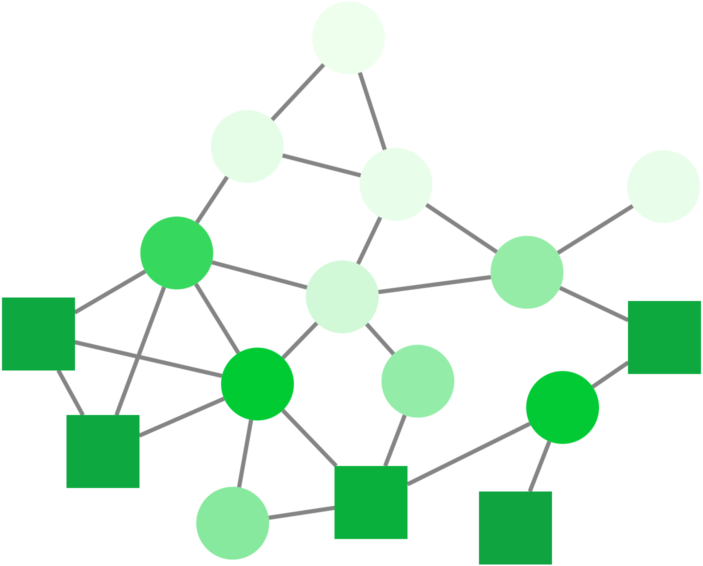
Network Propagation



Human proteins that interact with SARS-CoV-2 proteins

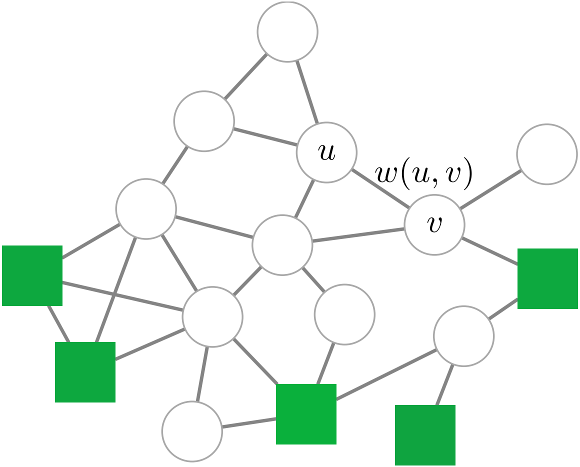
Other human proteins

Network Propagation



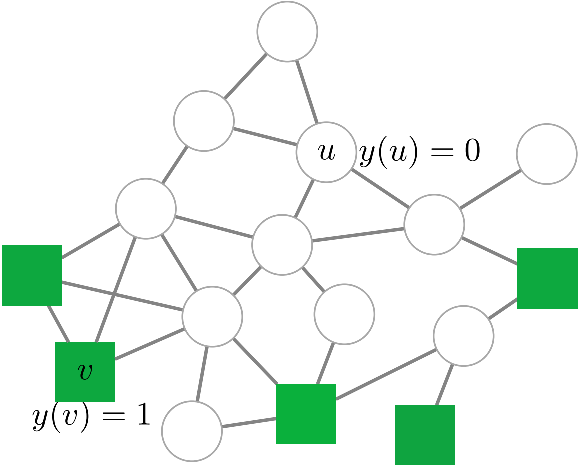
Cowen et al. Network propagation: a universal amplifier of genetic associations. Nat. Review Gen., 2017

Network Propagation: Regularized Laplacian (RL)



$G = (V, E, w)$: undirected, weighted network of human proteins

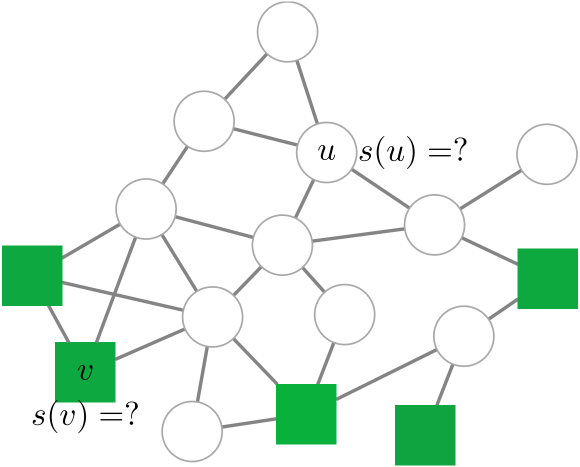
Network Propagation: Regularized Laplacian (RL)



y: label vector

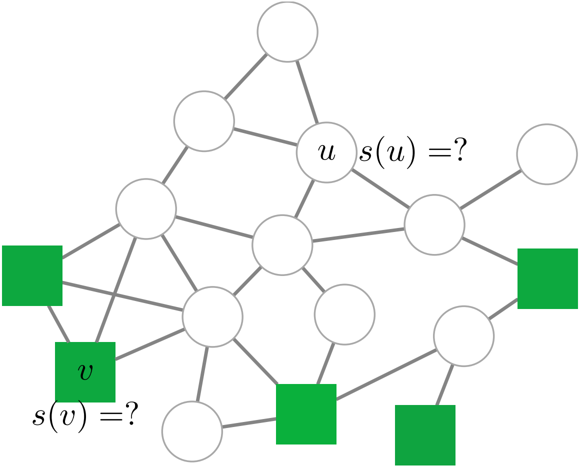
$$y(u) = \begin{cases} 1 & \text{if } u \text{ is a human interactor of virus (positive example)} \\ 0 & \text{otherwise} \end{cases}$$

Network Propagation: Regularized Laplacian (RL)



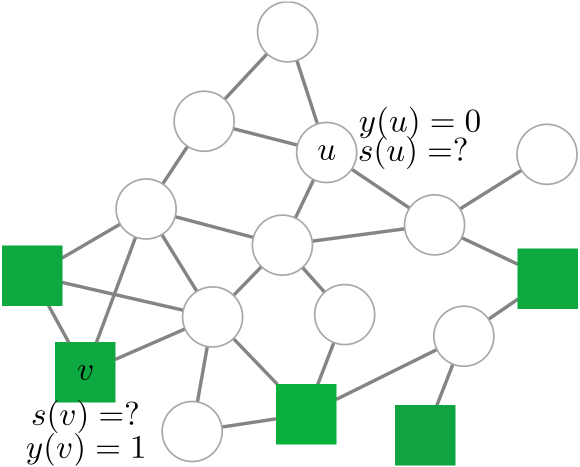
s: scores to compute

Network Propagation: Regularized Laplacian (RL)



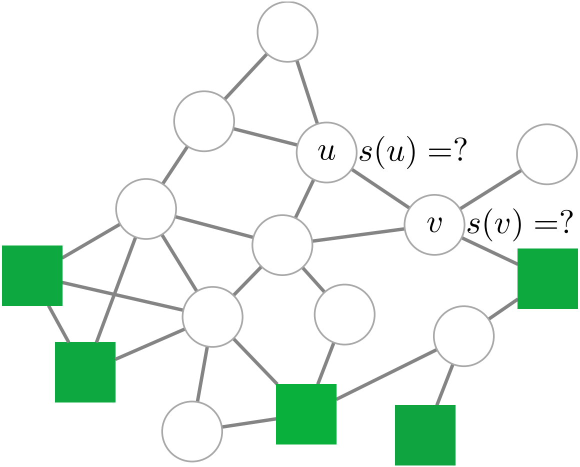
Minimize $\sum_{u \in V} \big(s(u) - y(u)\big)^2 + \alpha \sum_{(u,v) \in E} w_{uv}\big(s(u) - s(v)\big)^2$

Network Propagation: Regularized Laplacian (RL)



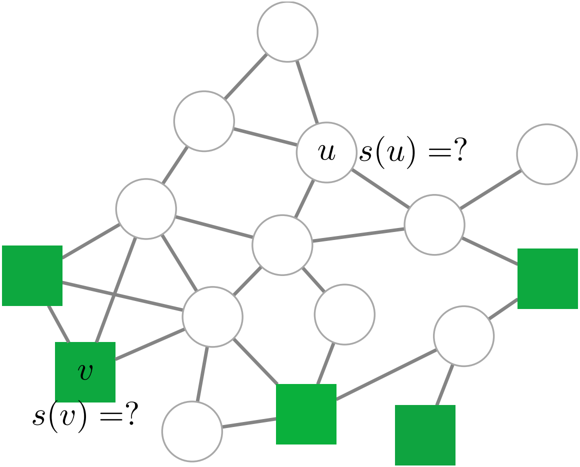
$$\text{Minimize} \sum_{u \in V} \left(s(u) - y(u)\right)^2 + \alpha \sum_{(u,v) \in E} w_{uv} \left(s(u) - s(v)\right)^2$$

Network Propagation: Regularized Laplacian (RL)



Minimize $\sum_{u \in V} \big(s(u) - y(u)\big)^2 + \alpha \sum_{(u,v) \in E} w_{uv}\big(s(u) - s(v)\big)^2$

Network Propagation: Regularized Laplacian (RL)

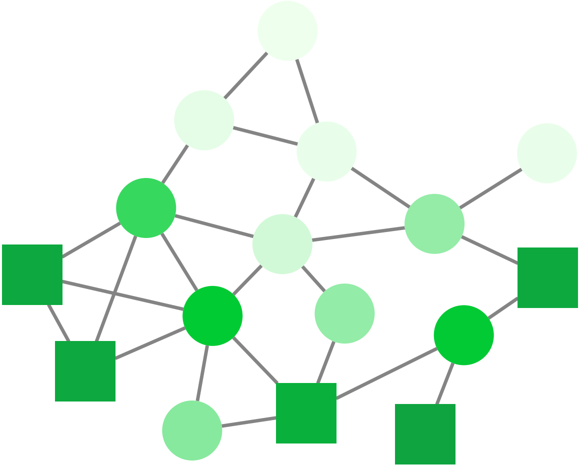


$$\text{Minimize} \sum_{u \in V} \big(s(u) - y(u)\big)^2 + \alpha \sum_{(u,v) \in E} w_{uv}\big(s(u) - s(v)\big)^2$$

$\alpha = 0$: set $s(u) = y(u)$

α very large: set $s(u) = s(v)$

Network Propagation: Regularized Laplacian (RL)



Solve linear system of equations: $s(v) = \dfrac{\alpha \sum_{u \in N(v)} w(u,v)s(u) + y(v)}{\alpha d(v) + 1}$

Matrix Formulation

$$s(v) = \frac{\alpha \sum_{u \in N(v)} w(u,v)s(u) + y(v)}{\alpha d(v) + 1}$$

$$s(v) + \alpha\left(d(v)s(v) - \sum_{u \in N(v)} w(u,v)s(u)\right) = y(v)$$

Matrix Formulation

$$s(v) = \frac{\alpha \sum_{u \in N(v)} w(u,v)s(u) + y(v)}{\alpha d(v) + 1}$$

$$s(v) + \alpha\left(d(v)s(v) - \sum_{u \in N(v)} w(u,v)s(u)\right) = y(v)$$

- W: adjacency matrix of G
- D: diagonal matrix where $D_{uu} = \sum_v w_{uv}$, for every node u in G
- $\tilde{W} = D^{-1/2}WD^{-1/2}$: normalized adjacency matrix of G
- $\tilde{L} = D - \tilde{W}$: *Laplacian* of G

Matrix Formulation

$$s(v) = \frac{\alpha \sum_{u \in N(v)} w(u,v)s(u) + y(v)}{\alpha d(v) + 1}$$

$$s(v) + \alpha\left(d(v)s(v) - \sum_{u \in N(v)} w(u,v)s(u)\right) = y(v)$$

- W: adjacency matrix of G
- D: diagonal matrix where $D_{uu} = \sum_v w_{uv}$, for every node u in G
- $\tilde{W} = D^{-1/2} W D^{-1/2}$: normalized adjacency matrix of G
- $\tilde{L} = D - \tilde{W}$: *Laplacian* of G

$$(I + \alpha\tilde{L})\mathbf{s} = \mathbf{y}$$

$$\mathbf{s} = (I + \alpha\tilde{L})^{-1}\mathbf{y}$$

Matrix Formulation

$$s(v) = \frac{\alpha \sum_{u \in N(v)} w(u,v)s(u) + y(v)}{\alpha d(v) + 1}$$

$$s(v) + \alpha\left(d(v)s(v) - \sum_{u \in N(v)} w(u,v)s(u)\right) = y(v)$$

- W: adjacency matrix of G
- D: diagonal matrix where $D_{uu} = \sum_v w_{uv}$, for every node u in G
- $\tilde{W} = D^{-1/2}WD^{-1/2}$: normalized adjacency matrix of G
- $\tilde{L} = D - \tilde{W}$: *Laplacian* of G

$$(I + \alpha\tilde{L})\mathbf{s} = \mathbf{y}$$

$$\mathbf{s} = (I + \alpha\tilde{L})^{-1}\mathbf{y}$$

- $(I + \alpha\tilde{L})^{-1}$ is the *regularized Laplacian*.
- Inverse exists if G is connected.
- Each entry records propagation between pair of nodes.

Algorithms

Network propagation:

- Regularized Laplacian[1] (RL)
- GeneMANIA[2]: RL with negative examples.
- SinkSource[3]: Like RL but fixes score of positive examples.

Supervised classification:

- Linear SVM, Logistic Regression
    - Feature vector = adjacency vector
    - L2 regularization

Deep learning:

- deepNF[4]: Autoencoder + Linear SVM

Guilt-by-association:

- Local: weighted average of neighbors

---

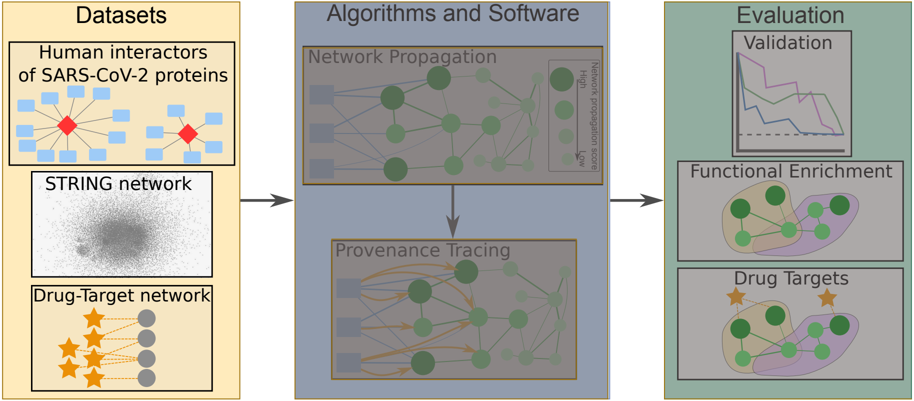
[1] *D. Zhou, B. Scholkopf. A regularization framework for learning from graph data. ICML Workshop 2004*

[2] *Mostafavi et al. GeneMANIA: a real-time multiple association network [...] Genome Biology 2008*

[3] *Murali et al. Network-based prediction and analysis of HIV dependency factors. PLoS Comput Biol 2011*

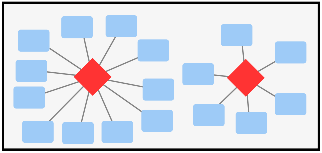
[4] *Gligorijević, Barot, and Bonneau, deepNF: deep network fusion for protein function prediction. Bioinformatics, 2018*

Datasets


Datasets
Human interactors of SARS-CoV-2 proteins
STRING network
Drug-Target network
Algorithms and Software
Network Propagation
Network propagation score
High
Low
Provenance Tracing
Evaluation
Validation
Functional Enrichment
Drug Targets

Datasets

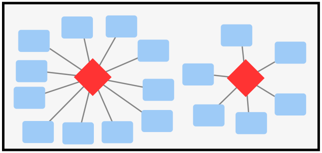
- SARS-CoV-2 interactors[1]
  - 332 Human interactors, 26 SARS-CoV-2 proteins



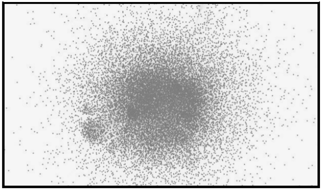
[1] Gordon et al. *A SARS-CoV-2-Human Protein-Protein Interaction Map [...]. Nature 2020*

Datasets

- SARS-CoV-2 interactors[1]
  - 332 Human interactors, 26 SARS-CoV-2 proteins



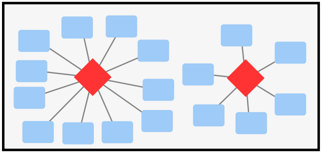
- Network of human proteins (STRING[2] version 11)
  - Interactions are "universal": independent of SARS-CoV-2
  - 19K nodes and 1M edges
  - 328 human interactors in network



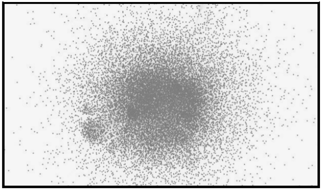
---

1. *Gordon et al. A SARS-CoV-2-Human Protein-Protein Interaction Map [...]. Nature 2020*
2. *Szklarczyk et al. STRING v11: protein-protein association networks [...]. Nucleic Acids Res 2019*

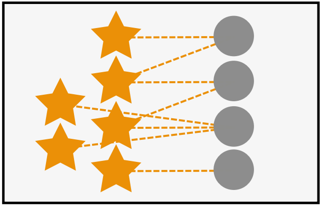
Datasets



- SARS-CoV-2 interactors[1]
  - 332 Human interactors, 26 SARS-CoV-2 proteins



- Network of human proteins (STRING[2] version 11)
  - Interactions are "universal": independent of SARS-CoV-2
  - 19K nodes and 1M edges
  - 328 human interactors in network

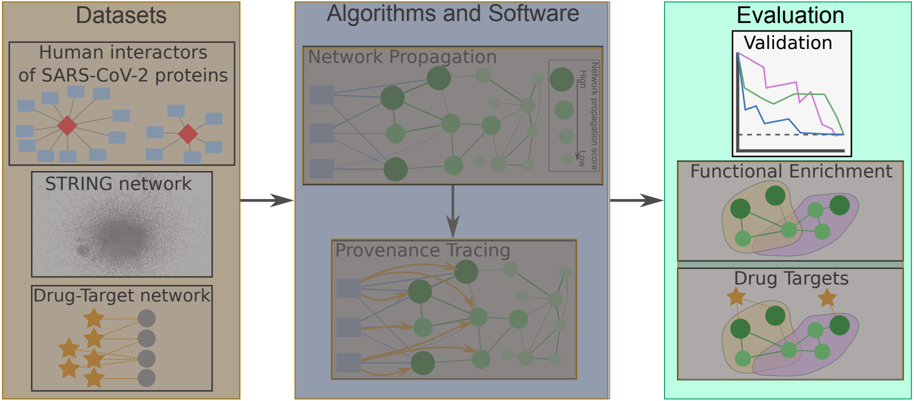


- Drug-Target dataset (DrugBank[3] version 5.1.6)
  - 6K drugs and 3K target (human) proteins
  - 16K drug-protein target pairs

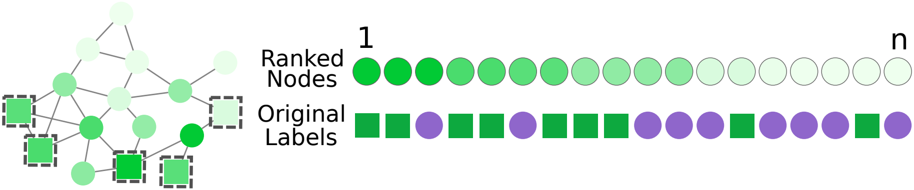
---

1. *Gordon et al. A SARS-CoV-2-Human Protein-Protein Interaction Map [...]. Nature 2020*
2. *Szklarczyk et al. STRING v11: protein-protein association networks [...]. Nucleic Acids Res 2019*
3. *Wishart et al. DrugBank 5.0: a major update to the DrugBank [...]. Nucleic Acids Res 2017*

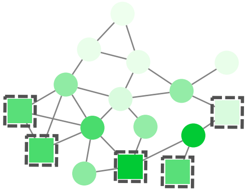
Evaluations

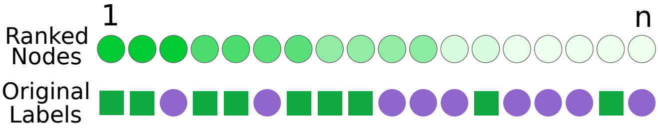


Five-Fold Cross-Validation



Five-Fold Cross-Validation



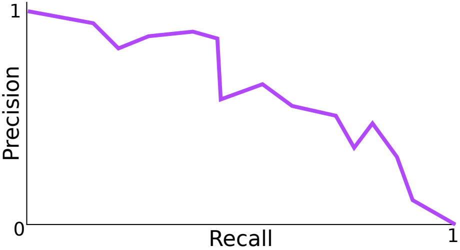


Recall:

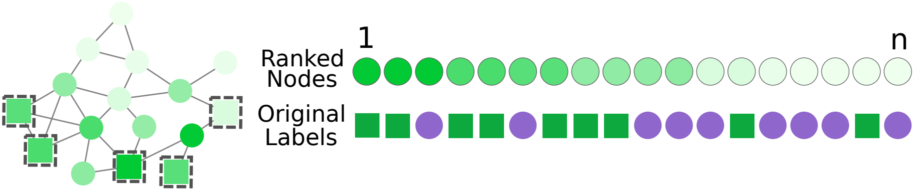
$$r_i = \frac{\text{true positives up to } i}{|P|}$$

Precision:

$$p_i = \frac{\text{true positives up to } i}{i}$$



Five-Fold Cross-Validation



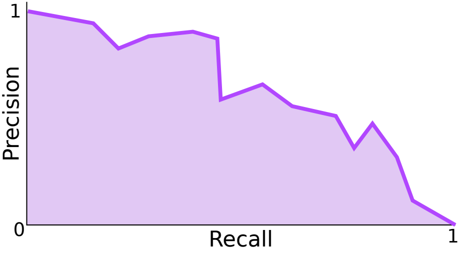
Recall:

$$r_i = \frac{\text{true positives up to } i}{|P|}$$

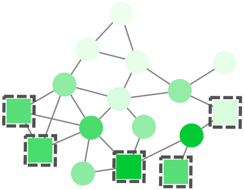
Precision:

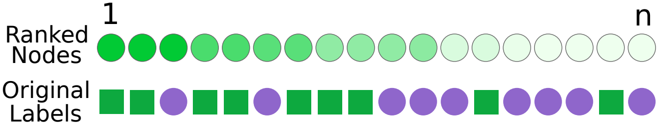
$$p_i = \frac{\text{true positives up to } i}{i}$$

- Area Under P-R Curve (AUPRC)



Five-Fold Cross-Validation



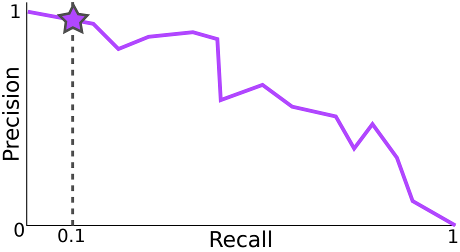


Recall:

$$r_i = \frac{\text{true positives up to } i}{|P|}$$

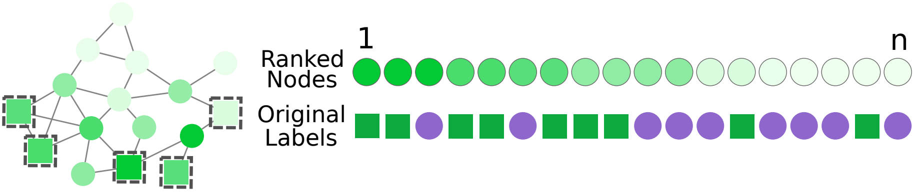
Precision:

$$p_i = \frac{\text{true positives up to } i}{i}$$



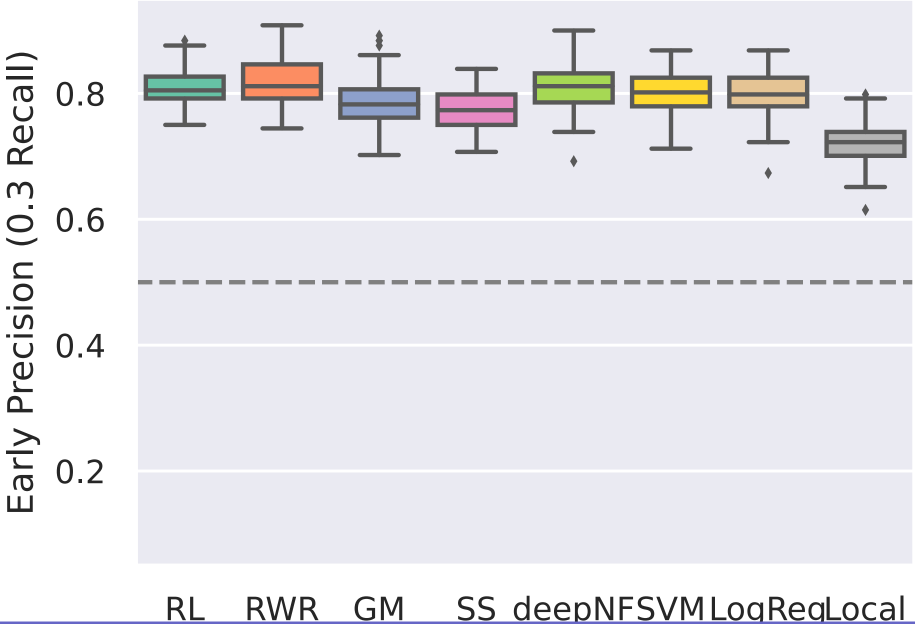
- Area Under P-R Curve (AUPRC)
- Early Precision (recall = 0.1)

Negative Examples

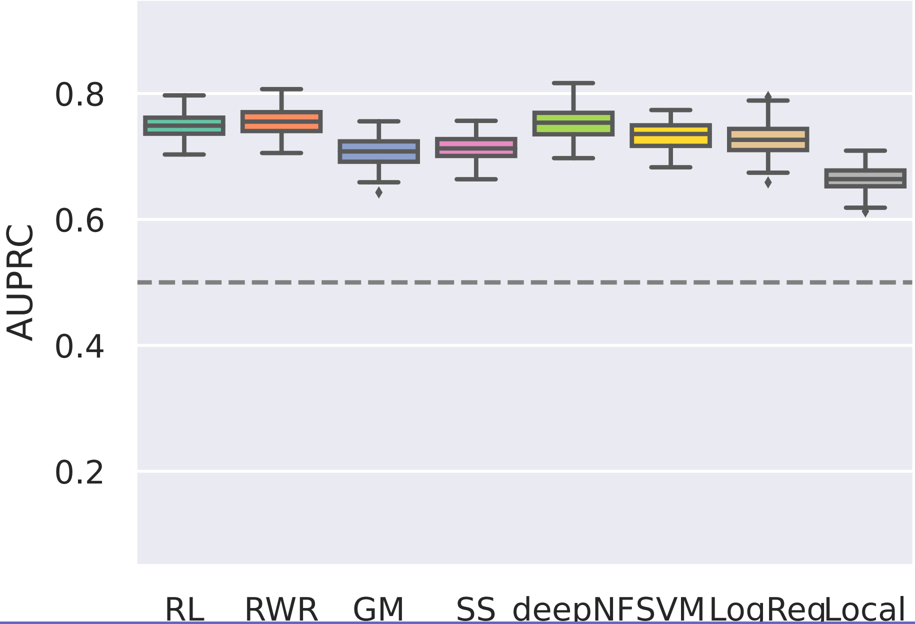


1. Required by methods
   - GeneMANIA, SVM, Logistic Regression
   - Averaged scores over 100 sets of randomly sampled negative examples
2. Needed to evaluate predictions
   - Early Precision, Area Under the Precision Recall Curve (AUPRC)
   - Repeated cross-validation 100 times
   - Positive:Negative (P:N) ratios 1:1, 1:5, and 1:10

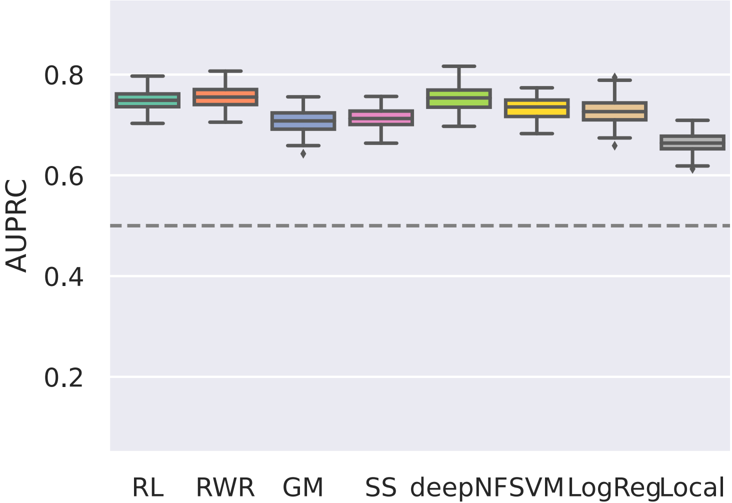
Five-Fold Cross-Validation (P:N ratio 1:1)



Five-Fold Cross-Validation (P:N ratio 1:1)

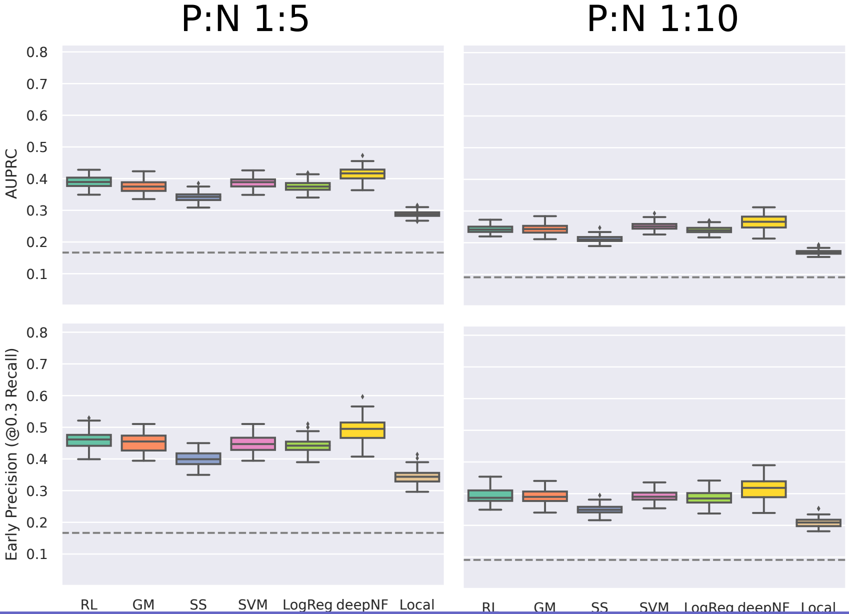


Five-Fold Cross-Validation (P:N ratio 1:1)

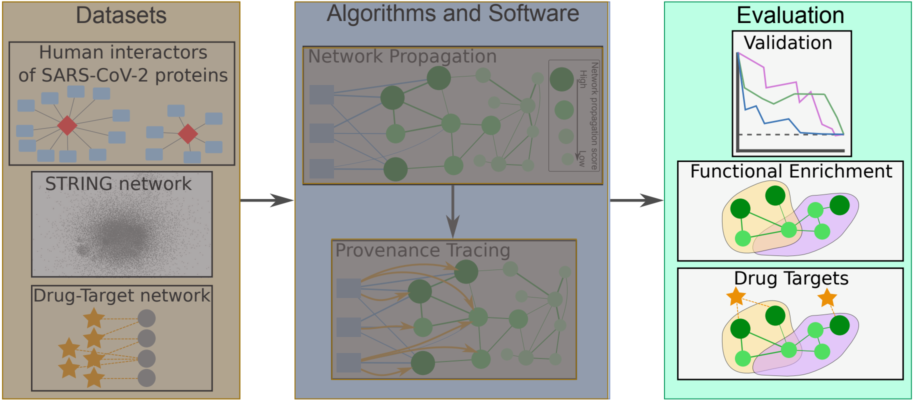


Network propagation is a promising approach to predict human proteins that interact with SARS-CoV-2

Other Positive:Negative Ratios



Further Analysis



Prediction Analysis

- Stratified sampling for estimating statistical significance of node scores.
    - Compare node's score to distribution of 1,000 randomly selected nodes with same degree.
    - Retain nodes with p-value ≤ 0.05

- Top 332 predictions of RL

Prediction Analysis

- Stratified sampling for estimating statistical significance of node scores.
    - Compare node's score to distribution of 1,000 randomly selected nodes with same degree.
    - Retain nodes with p-value ≤ 0.05
- Top 332 predictions of RL
- Tested for significant overlap with experimental human-SARS-CoV-2 PPI datasets.
- Tested for enrichment of GO Biological Process (BP) terms
    - Benjamini-Hochberg-corrected p-value < 0.01
    - $180+$ terms for RL

Prediction Analysis

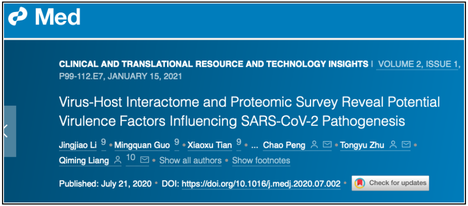
- Stratified sampling for estimating statistical significance of node scores.
    - Compare node's score to distribution of 1,000 randomly selected nodes with same degree.
    - Retain nodes with p-value ≤ 0.05

- Top 332 predictions of RL

- Tested for significant overlap with experimental human-SARS-CoV-2 PPI datasets.
- Tested for enrichment of GO Biological Process (BP) terms
    - Benjamini-Hochberg-corrected p-value < 0.01
    - $180+$ terms for RL

- Designed heuristic "Set-cover" algorithm to choose representative non-overlapping terms
    - 21 terms

New SARS-CoV-2-Human PPI Datasets



Med

CLINICAL AND TRANSLATIONAL RESOURCE AND TECHNOLOGY INSIGHTS | VOLUME 2, ISSUE 1, P99-112.E7, JANUARY 15, 2021

Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis

Jingjiao Li [9] • Mingquan Guo [9] • Xiaoxu Tian [9] • ... Chao Peng • Tongyu Zhu • Qiming Liang [10] • Show all authors • Show footnotes

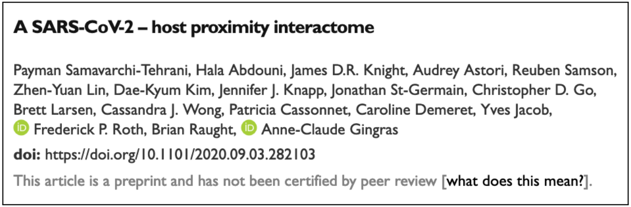
Published: July 21, 2020 • DOI: https://doi.org/10.1016/j.medj.2020.07.002 • Check for updates

Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV

Alexey Stukalov, Virginie Girault, Vincent Grass, Ozge Karayel, Valter Bergant, Christian Urban, Darya A. Haas, Yiqi Huang, Lila Oubraham, Anqi Wang, M. Sabri Hamad, Antonio Piras, Fynn M. Hansen, Maria C. Tanzer, Igor Paron, Luca Zinzula, Thomas Enghleitner, Maria Reinecke, Teresa M. Lavacca, Rosina Ehmann, Roman Wölfel, Jörg Jores, Bernhard Kuster, Ulrike Protzer, Roland Rad, John Ziebuhr, Volker Thiel, Pietro Scaturro, Matthias Mann, Andreas Pichlmair

doi: https://doi.org/10.1101/2020.06.17.156455

This article is a preprint and has not been certified by peer review [what does this mean?].



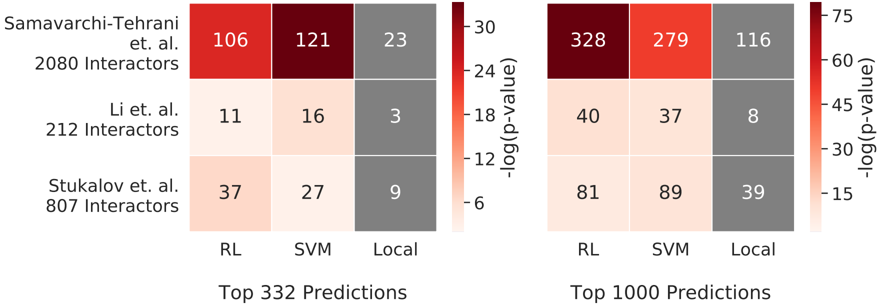
A SARS-CoV-2 – host proximity interactome

Payman Samavarchi-Tehrani, Hala Abdouni, James D.R. Knight, Audrey Astori, Reuben Samson, Zhen-Yuan Lin, Dae-Kyum Kim, Jennifer J. Knapp, Jonathan St-Germain, Christopher D. Go, Brett Larsen, Cassandra J. Wong, Patricia Cassonnet, Caroline Demeret, Yves Jacob, Frederick P. Roth, Brian Raught, Anne-Claude Gingras

doi: https://doi.org/10.1101/2020.09.03.282103

This article is a preprint and has not been certified by peer review [what does this mean?].

Overlap with New SARS-CoV-2-Human PPI Datasets



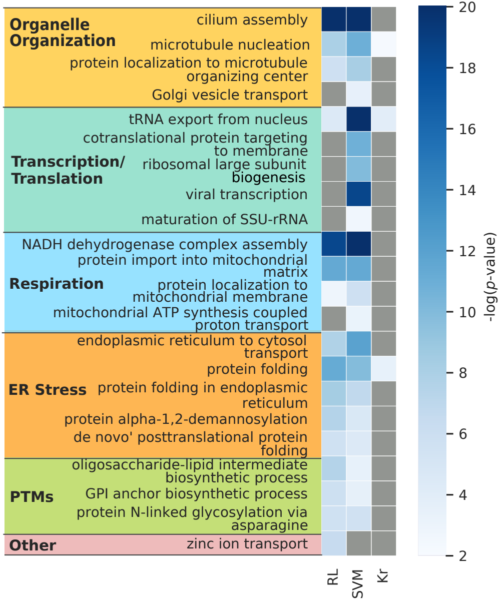
Top 332 Predictions

	RL	SVM	Local
Samavarchi-Tehrani et. al. 2080 Interactors	106	121	23
Li et. al. 212 Interactors	11	16	3
Stukalov et. al. 807 Interactors	37	27	9

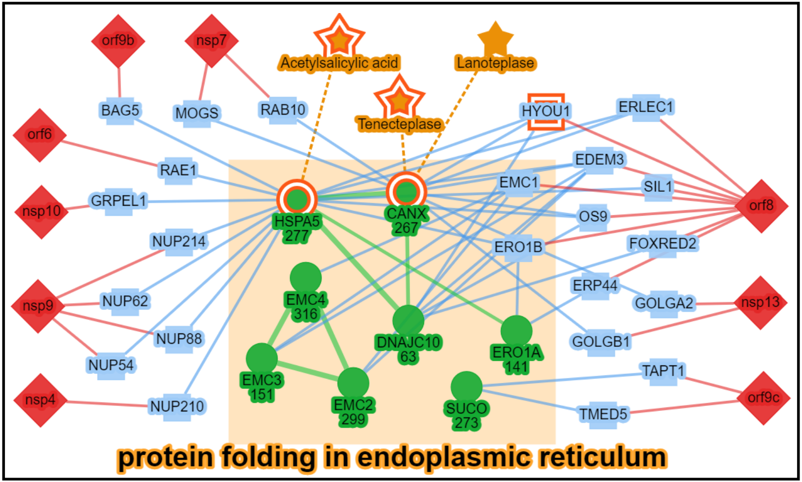
Top 1000 Predictions

	RL	SVM	Local
Samavarchi-Tehrani et. al. 2080 Interactors	328	279	116
Li et. al. 212 Interactors	40	37	8
Stukalov et. al. 807 Interactors	81	89	39

Functional Enrichment Summary



Functional Enrichment Summary



◆ SARS-CoV-2 protein

■ Human interactor of SARS-CoV-2

● Top-ranking human protein

★ Drug

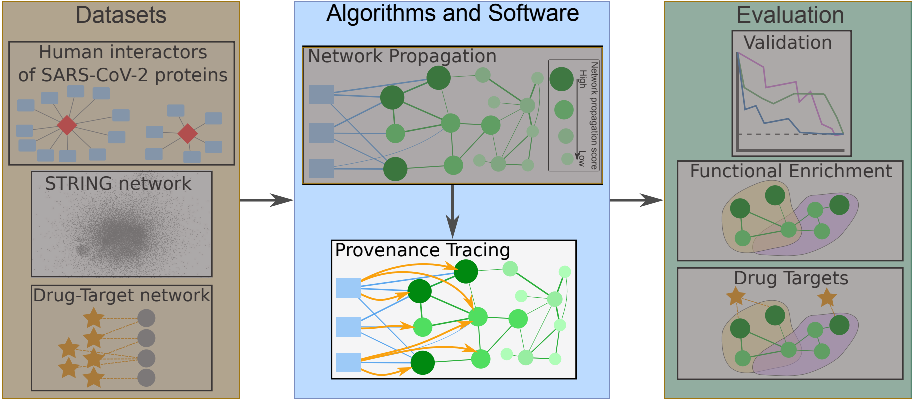
✪ Drug under COVID-19 trial

— SARS-CoV-2 - Human PPI

— Human PPI

- - - Protein-drug interaction

Interpetable Network Propagation



Datasets

Human interactors of SARS-CoV-2 proteins

STRING network

Drug-Target network

Algorithms and Software

Network Propagation

Provenance Tracing

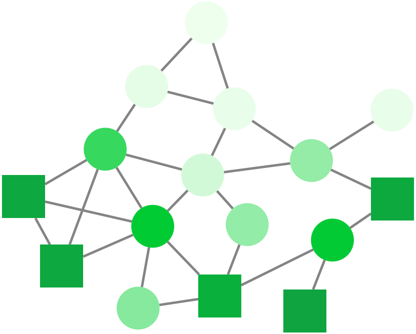
Evaluation

Validation

Functional Enrichment

Drug Targets

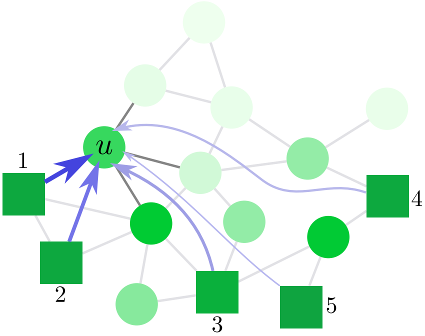
Provenance Tracing



- Compute a reproducible trace of every prediction back to experimental sources.

Kasif and Roberts, We need to keep a reproducible trace of facts, predictions, and hypotheses from gene to function in the era of big data, PLoS Biol., 2020.

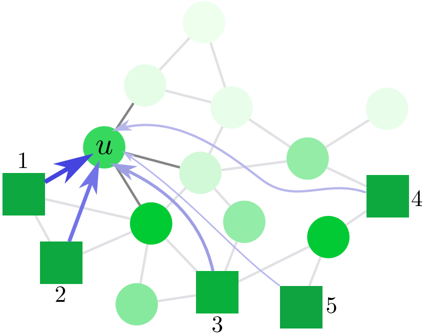
Provenance Tracing



- Compute a reproducible trace of every prediction back to experimental sources.
- For each node u with score $s(u)$, rank every positive example by the contribution it makes to $s(u)$.

Kasif and Roberts, We need to keep a reproducible trace of facts, predictions, and hypotheses from gene to function in the era of big data, PLoS Biol., 2020.

Provenance Tracing



$$s = (I + \alpha\tilde{L})^{-1}y$$

$$K = (I + \alpha\tilde{L})^{-1}$$

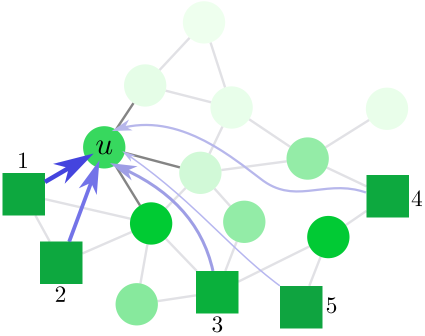
$$s(u) = \sum_{v \in P} K_{uv}$$

Every score $s(u)$ is a sum of contributions from SARS-CoV-2 interactors.

- Compute a reproducible trace of every prediction back to experimental sources.
- For each node u with score $s(u)$, rank every positive example by the contribution it makes to $s(u)$.

Kasif and Roberts, We need to keep a reproducible trace of facts, predictions, and hypotheses from gene to function in the era of big data, PLoS Biol., 2020.

Provenance Tracing



$$s = (I + \alpha\tilde{L})^{-1}y$$

$$K = (I + \alpha\tilde{L})^{-1}$$

$$s(u) = \sum_{v \in P} K_{uv}$$

Every score $s(u)$ is a sum of contributions from SARS-CoV-2 interactors.

- Compute a reproducible trace of every prediction back to experimental sources.
- For each node u with score $s(u)$, rank every positive example v in decreasing order of K_{uv}.
- General purpose strategy for large class of network propagation algorithms.

Kasif and Roberts, We need to keep a reproducible trace of facts, predictions, and hypotheses from gene to function in the era of big data, PLoS Biol., 2020.

Results of Provenance Tracing

- When α is small, expect highest contributing sources to be direct neighbors of top-ranking proteins.
- As α increases, expect more of the highest contributors to not be directly connected by an edge to top-ranking proteins.
- Evaluated six values of α between 0.01 and 100.

Results of Provenance Tracing

- When α is small, expect highest contributing sources to be direct neighbors of top-ranking proteins.
- As α increases, expect more of the highest contributors to not be directly connected by an edge to top-ranking proteins.
- Evaluated six values of α between 0.01 and 100.

- For every value of α and for every top-ranking protein u (till a rank of 1,000), the source protein with the highest contribution to $s(u)$ was always a neighbor of u.
- Second and third highest contributors were more than one edge away for as few as 2% of the top-ranking proteins for $\alpha = 0.01$.
- This number increased only to 25% for $\alpha = 100$.

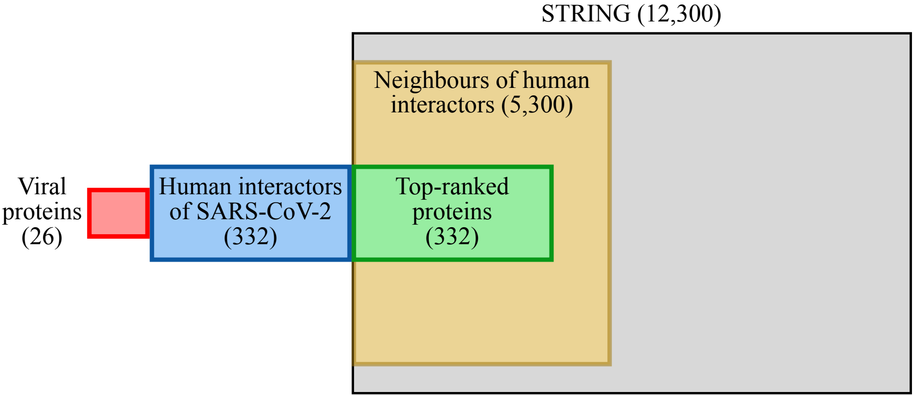
Results of Provenance Tracing

- When α is small, expect highest contributing sources to be direct neighbors of top-ranking proteins.
- As α increases, expect more of the highest contributors to not be directly connected by an edge to top-ranking proteins.
- Evaluated six values of α between 0.01 and 100.

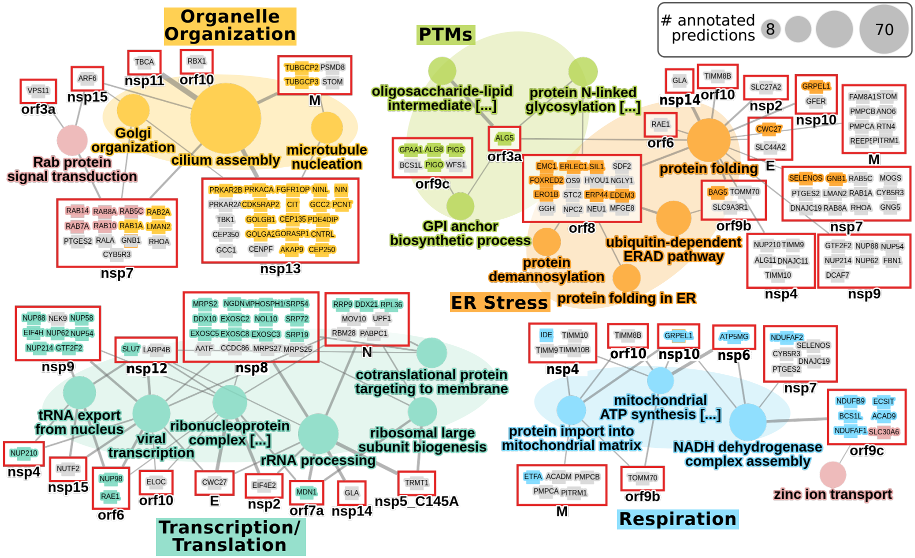
- For every value of α and for every top-ranking protein u (till a rank of 1,000), the source protein with the highest contribution to $s(u)$ was always a neighbor of u.
- Second and third highest contributors were more than one edge away for as few as 2% of the top-ranking proteins for $\alpha = 0.01$.
- This number increased only to 25% for $\alpha = 100$.

- 332 positive examples have 5,300 neighbours (out of 12,300 proteins in STRING).

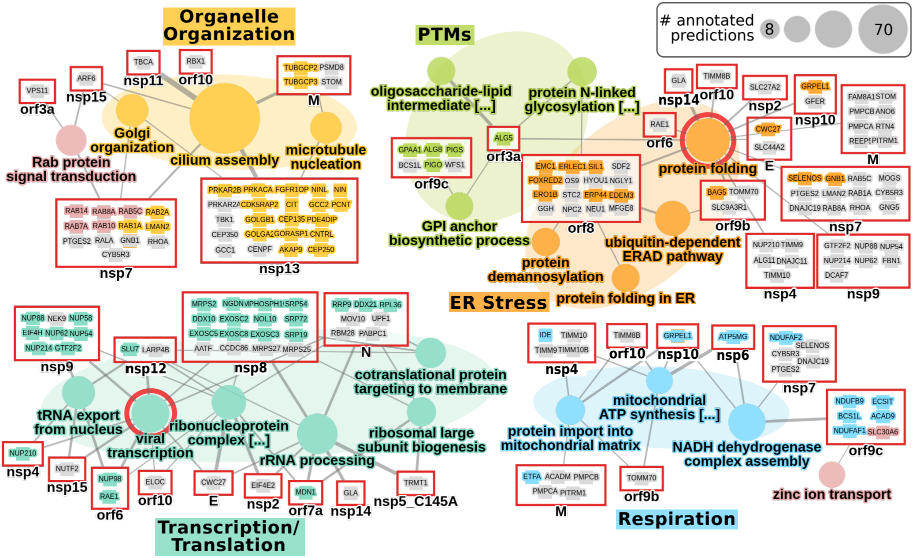
Understanding Provenance Tracing Results



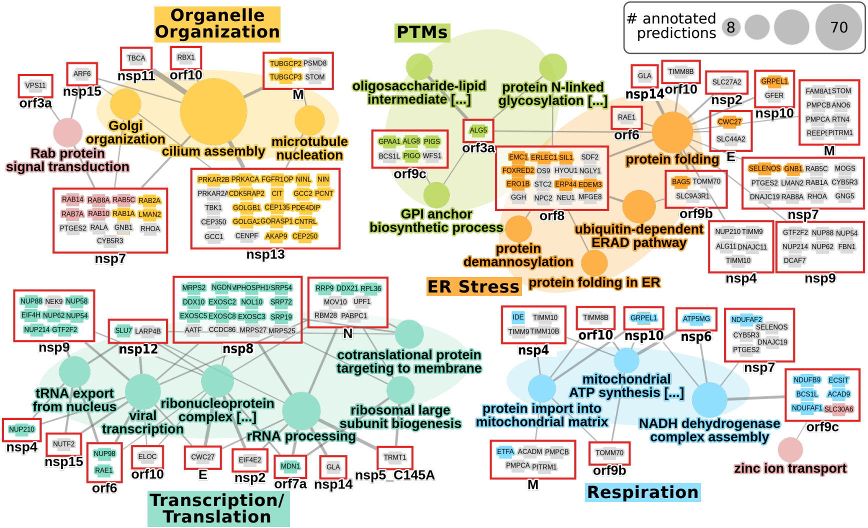
GO Term - Human-Viral Interactors Overview



GO Term - Human-Viral Interactors Overview

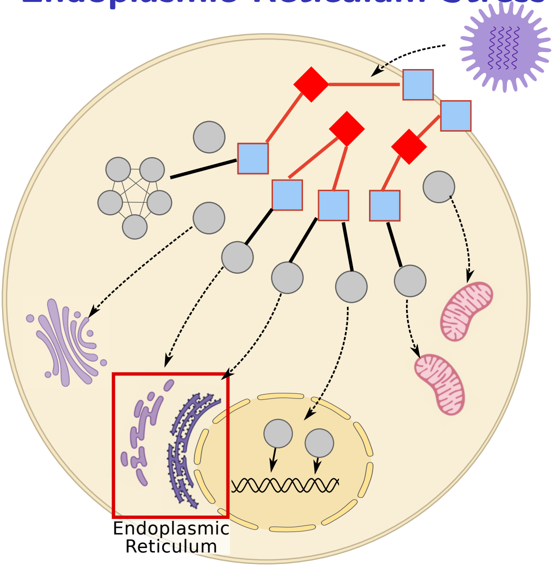


GO Term - Human-Viral Interactors Overview



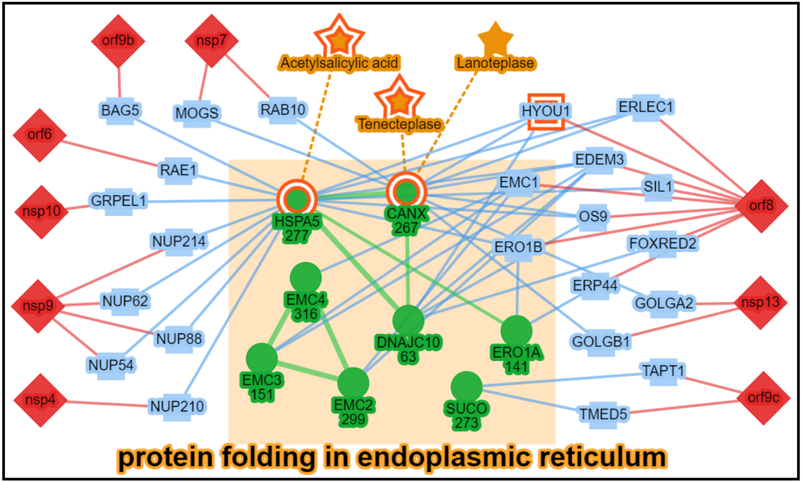
Literature support exists for most terms (SARS-CoV-2 or closely related virus)

Endoplasmic Reticulum Stress



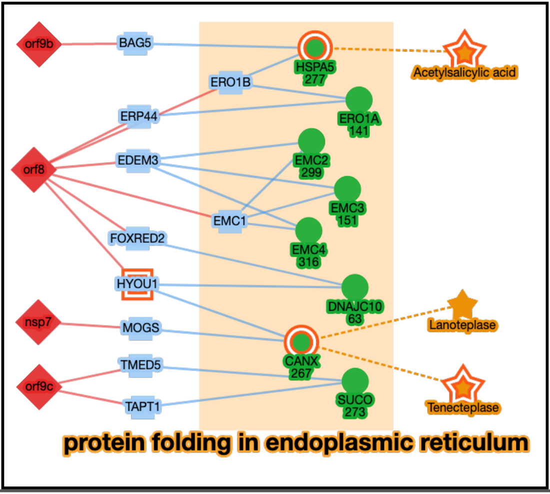
Chan et al. Modulation of the Unfolded Protein Response by the [SARS-CoV] Spike Protein. J. of Virology 2006
Diego et al. [SARS-CoV] Envelope Protein Regulates Cell Stress Response and Apoptosis. PLoS Pathogens 2011
Koseler et al. Endoplasmic Reticulum Stress Markers in SARS-CoV-2 Infection [...] In Vivo 2020

Protein Folding in ER



- 🟥 SARS-CoV-2 protein
- 🟦 Human interactor of SARS-CoV-2
- ● Top-ranking human protein
- ★ Drug
- ★ Drug under COVID-19 trial
- SARS-CoV-2 - Human PPI
- Human PPI
- Protein-drug interaction

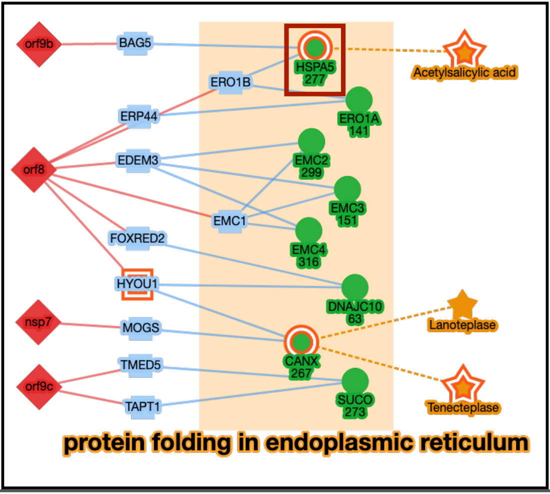
Protein Folding in ER



◆ SARS-CoV-2 protein
■ Human interactor of SARS-CoV-2
● Top-ranking human protein
★ Drug
☆ Drug under COVID-19 trial

— SARS-CoV-2 - Human PPI
— Human PPI
- - - Protein-drug interaction

Protein Folding in ER



◆ SARS-CoV-2 protein

■ Human interactor of SARS-CoV-2

● Top-ranking human protein

★ Drug

☆ Drug under COVID-19 trial

— SARS-CoV-2 - Human PPI

— Human PPI

- - - Protein-drug interaction

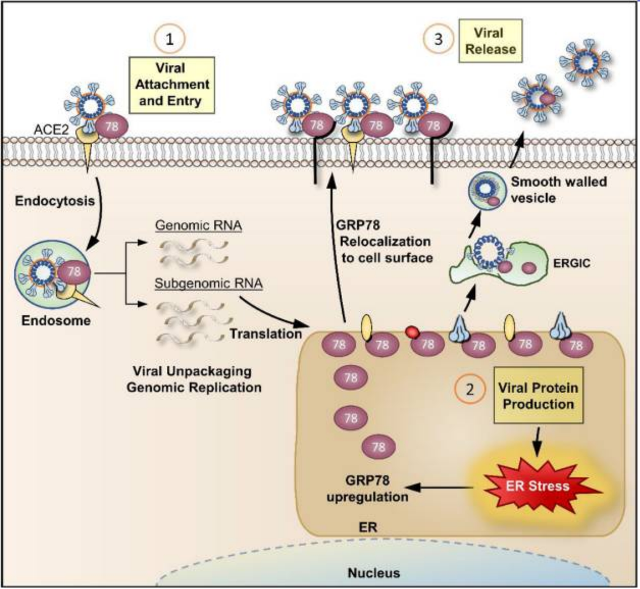
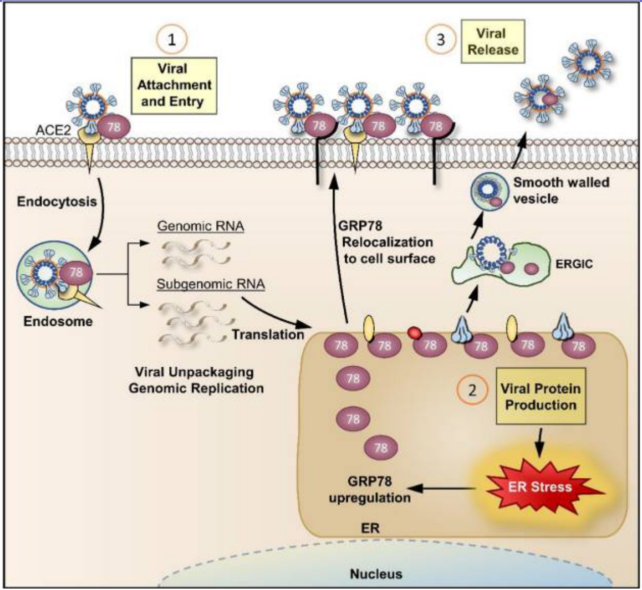


Figure: Ha et al. The stress-inducible molecular chaperone GRP78 as potential [...]. J. Infection 2020
1. Chu et al. [MERS] and bat coronavirus HKU9 both can utilize GRP78 for attachment [...]. J.B.C. 2018
2. Chan et al. Modulation of the Unfolded Protein Response by the [SARS-CoV] Spike Protein. J. Virology 2006
3. Wu et al. Japanese encephalitis virus co-opts the ER-stress response protein GRP78 [...]. Virology J 2011



Predicted receptor for SARS-CoV-2 Spike protein[4]

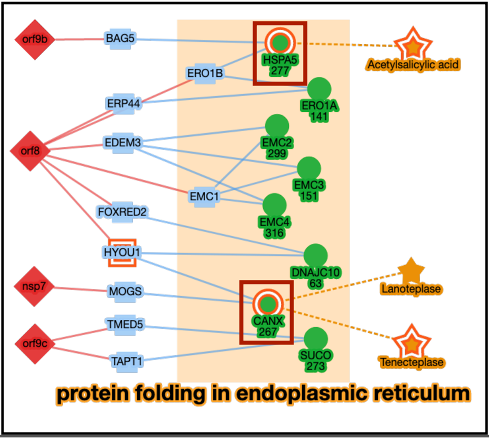
WEBMD HEALTH NEWS

Blood Clots Are Another Dangerous COVID-19 Mystery

By Brenda Goodman, MA

Terpos et al. Hematological findings and complications of COVID-19 Am. J. Hematol. 2020
Helms et al. High risk of thrombosis in patients with severe SARS-CoV-2 infection [...] Intensive Care Med 2020

Protein Folding in ER: Connection to Blood Clotting

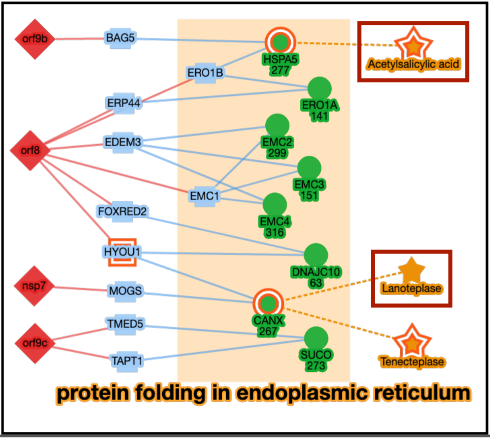


- 🟥 SARS-CoV-2 protein
- 🟦 Human interactor of SARS-CoV-2
- ● Top-ranking human protein
- ★ Drug
- ☆ Drug under COVID-19 trial
- — SARS-CoV-2 - Human PPI
- — Human PPI
- - - Protein-drug interaction

HSPA5 and CANX are chaperones for pro-coagulant proteins

Williams et al. Beyond Lectins: The calnexin/calreticulin Chaperone System of the [ER] J. Cell. Sci 2006
Tang et al. Anticoagulant treatment is associated with decreased mortality [COVID-19] [...] J. Thromb H. 2020

Protein Folding in ER: Connection to Blood Clotting



◆ SARS-CoV-2 protein
■ Human interactor of SARS-CoV-2
● Top-ranking human protein
★ Drug
✩ Drug under COVID-19 trial
— SARS-CoV-2 - Human PPI
— Human PPI
--- Protein-drug interaction

Clinical trials: Aspirin and Alteplase (derivative of Lanoteplase)

Williams et al. Beyond Lectins: The calnexin/calreticulin Chaperone System of the [ER] J. Cell. Sci 2006
Tang et al. Anticoagulant treatment is associated with decreased mortality [COVID-19] [...] J. Thromb H. 2020

Summary

- Network propagation well-suited to pinpoint processes that may be hijacked by the virus

Summary

- Network propagation well-suited to pinpoint processes that may be hijacked by the virus

- Prioritized list of human proteins and drug targets
  - Predictions supported by literature
  - 6+ drugs already in clinical trials

Summary

- Network propagation well-suited to pinpoint processes that may be hijacked by the virus
- Prioritized list of human proteins and drug targets
  - Predictions supported by literature
  - 6+ drugs already in clinical trials
- Promising direction to target ER stress, HSPA5

Summary

- Network propagation well-suited to pinpoint processes that may be hijacked by the virus
- Prioritized list of human proteins and drug targets
    - Predictions supported by literature
    - 6+ drugs already in clinical trials
- Promising direction to target ER stress, HSPA5
- Additional promising target: Cilium assembly and Tubulins

Summary

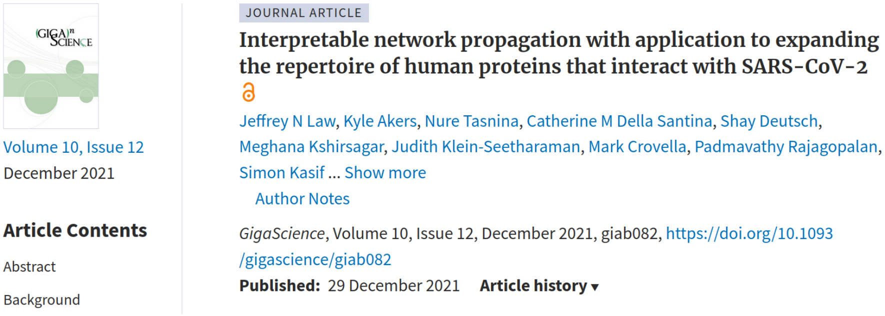
- Network propagation well-suited to pinpoint processes that may be hijacked by the virus
- Prioritized list of human proteins and drug targets
  - Predictions supported by literature
  - 6+ drugs already in clinical trials
- Promising direction to target ER stress, HSPA5
- Additional promising target: Cilium assembly and Tubulins
- Substantial overlap new experimental datasets.

Samavarchi-Tehrani et al., "A SARS-CoV-2 – host proximity interactome," bioRxiv, September 4, 2020.

Summary

- Network propagation well-suited to pinpoint processes that may be hijacked by the virus
- Prioritized list of human proteins and drug targets
    - Predictions supported by literature
    - 6+ drugs already in clinical trials
- Promising direction to target ER stress, HSPA5
- Additional promising target: Cilium assembly and Tubulins
- Substantial overlap new experimental datasets.
- New method for provenace tracing for large class of network propagation algorithms

Availability



Volume 10, Issue 12
December 2021

Article Contents

Abstract

Background

JOURNAL ARTICLE

Interpretable network propagation with application to expanding the repertoire of human proteins that interact with SARS-CoV-2

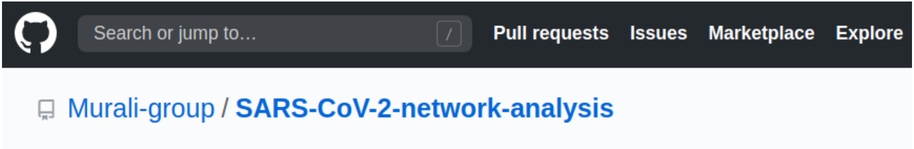
Jeffrey N Law, Kyle Akers, Nure Tasnina, Catherine M Della Santina, Shay Deutsch, Meghana Kshirsagar, Judith Klein-Seetharaman, Mark Crovella, Padmavathy Rajagopalan, Simon Kasif ... Show more

Author Notes

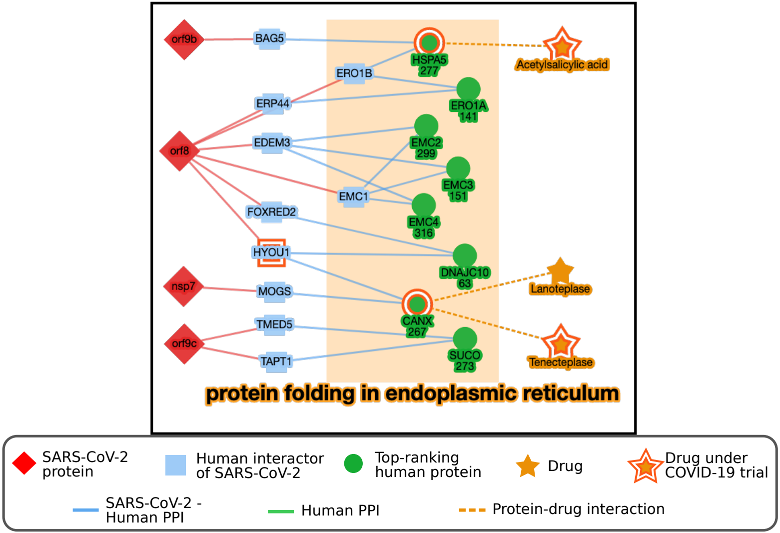
GigaScience, Volume 10, Issue 12, December 2021, giab082, https://doi.org/10.1093/gigascience/giab082

Published: 29 December 2021 **Article history ▾**

`https://doi.org/10.1093/gigascience/giab082`



Murali-group / **SARS-CoV-2-network-analysis**



Available on GraphSpace: https://bit.ly/2Ap268Z

Acknowledgements

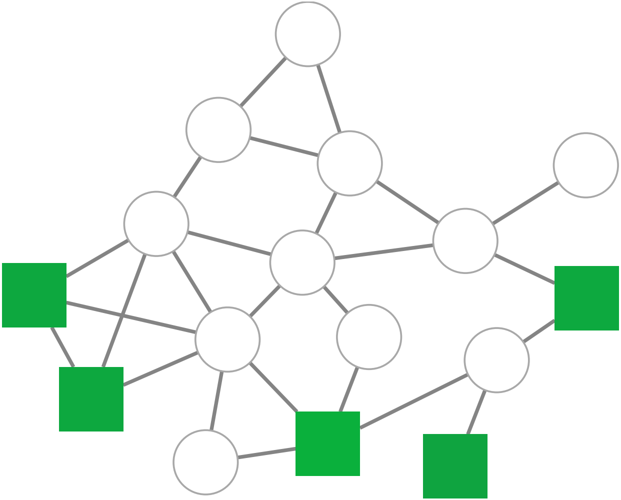
Jeffrey N. Law[1], Kyle Akers[1], Nure Tasnina[2], Catherine M. Della Santina[3],
Meghana Kshirsagar[4], Mark Crovella[5], Padmavathy Rajagopalan[6],
Judith Klein-Seetharaman[7], Simon Kasif[2,3]

1. Genetics, Bioinformatics, and Comp. Biology Ph.D. program, Virginia Tech
2. Department of Computer Science, Virginia Tech,
3. Department of Biomedical Engineering, Boston University
4. AI for Good Lab, Microsoft
5. Department of Computer Science, Boston University
6. Department of Chemical Engineering, Virginia Tech
7. Department of Chemistry, Colorado School of Mines

Funding:

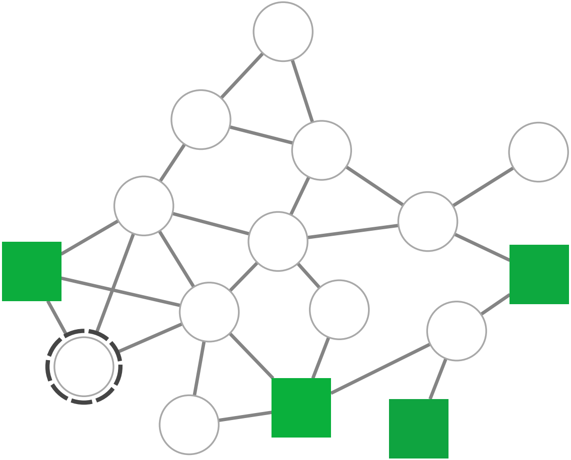
- NSF: DBI-1759858, MCB-1817736, CCF-2029543, CNS-1618207, CBET-1510920
- USDA-NIFA: 201807578
- Genetics, Bioinformatics, and Computational Biology program at Virginia Tech
- Computational Tissue Engineering Graduate Education Program at Virginia Tech
- Hariri Institute and the Department of Biomedical Engineering at Boston University

Validation: Leave-one-out Propagation Analysis



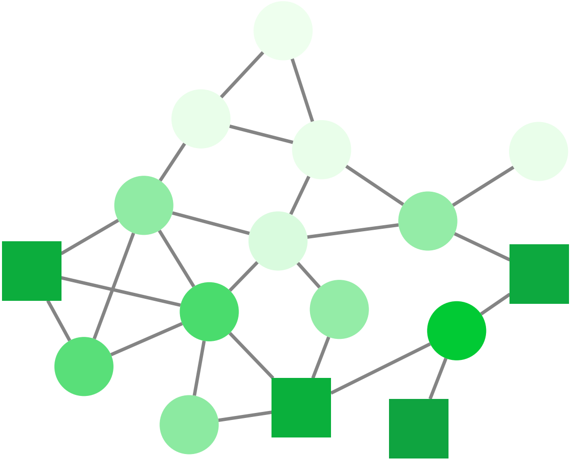
Validation: Leave-one-out Propagation Analysis

Leave out node



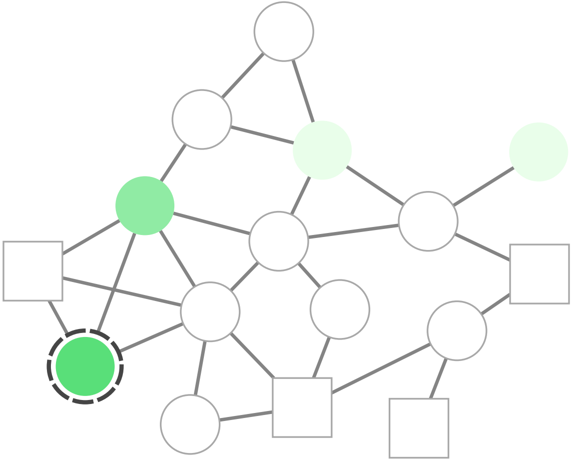
Validation: Leave-one-out Propagation Analysis

Network propagation

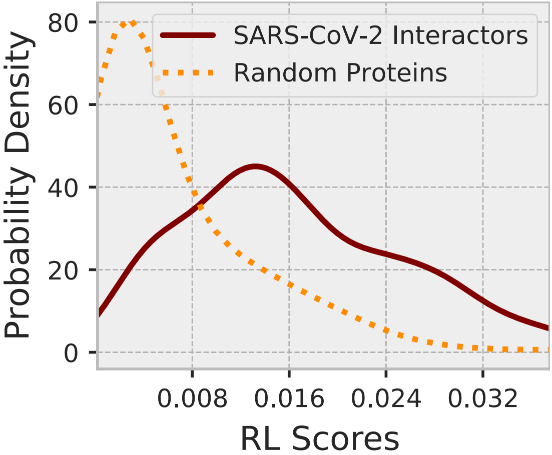


Validation: Leave-one-out Propagation Analysis

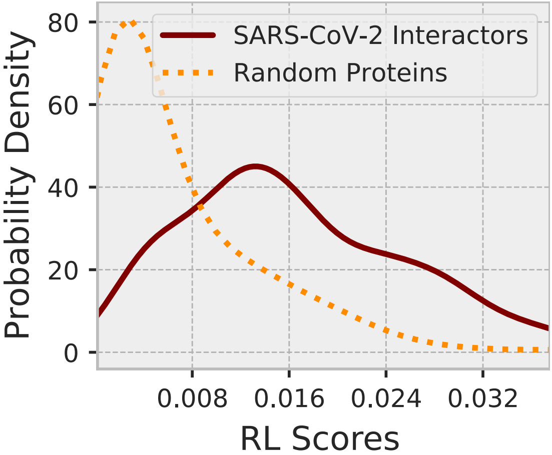
Evaluate vs random proteins



Validation: Leave-one-out Propagation Analysis



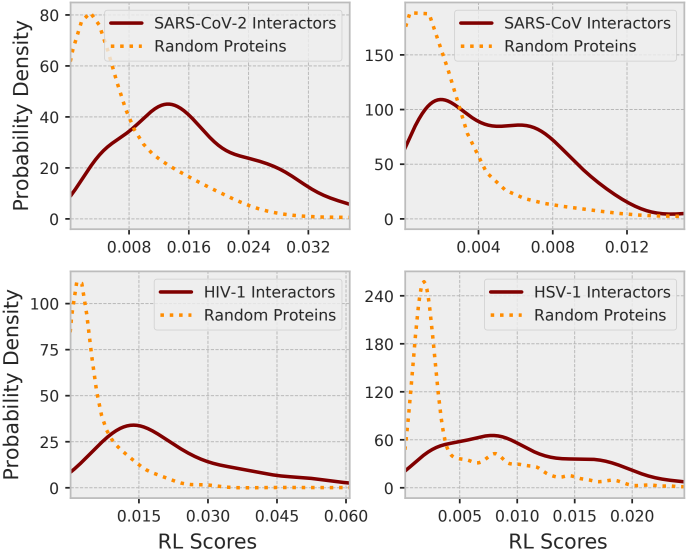
Validation: Leave-one-out Propagation Analysis



Kolmogorov-Smirnov test p-value 2.1×10^{-78}

Human interactors of virus are proximal in network

Validation: Leave-one-out Propagation Analysis



Kolmogorov-Smirnov test p-value $< 1 \times 10^{-27}$

Effect of Parameters