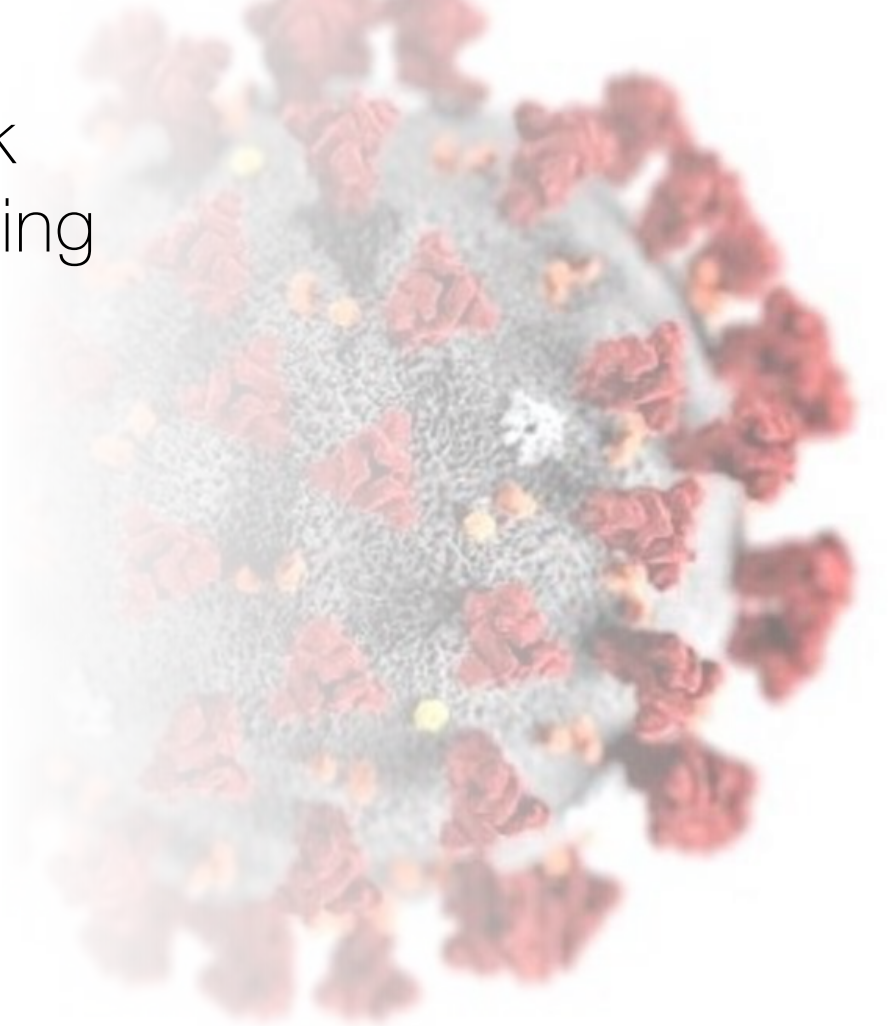
Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19



Deisy Morselli Gysi, Ítalo Do Valle,
Marinka Zitnik, Asher Ameli, Xiao Gan,
Onur Varol, Helia Sanchez, Rebecca Marlene Baron,
Dina Ghiassian, Joseph Loscalzo, Albert-László Barabási

NORTHEASTERN UNIVERSITY
HARVARD UNIVERSITY
CENTRAL EUROPEAN UNIVERSITY, BUDAPEST
SCIPHER MEDICINE

Manuscript: https://arxiv.org/abs/2004.07229

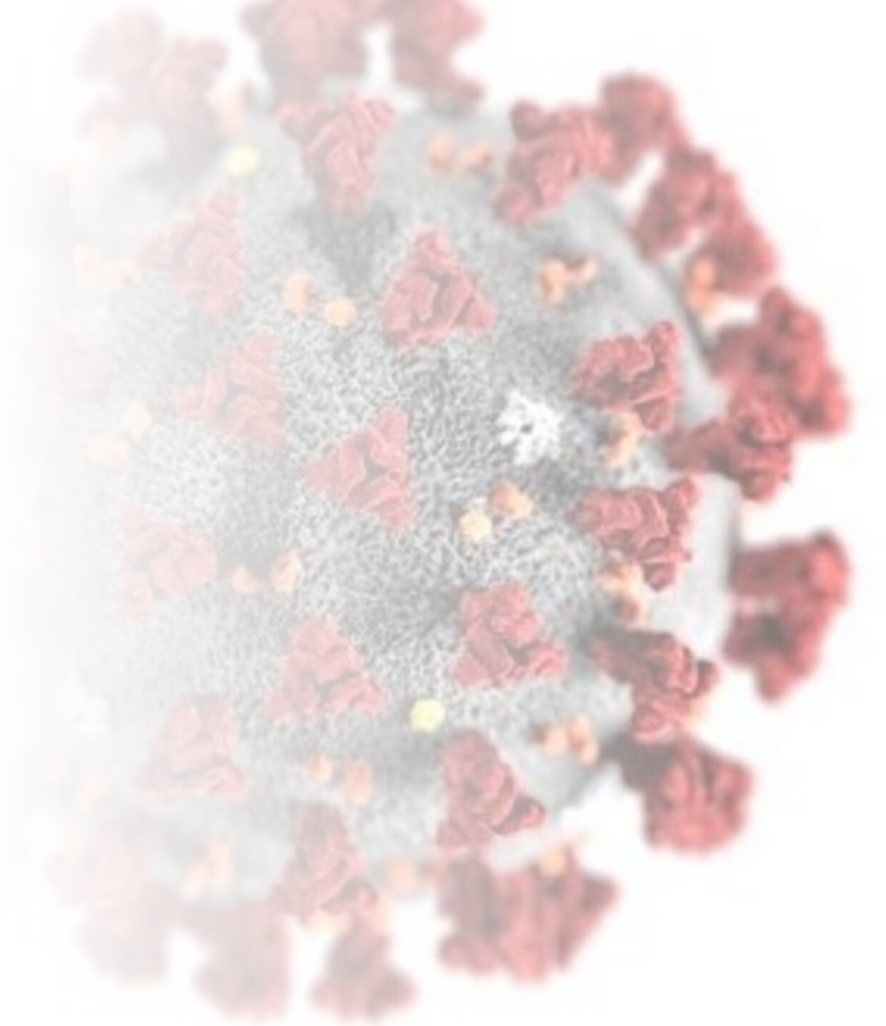
barabasi@gmail.com

The disruptive nature of the COVID-19 pandemic demands the rapid deployment of effective therapeutic interventions.

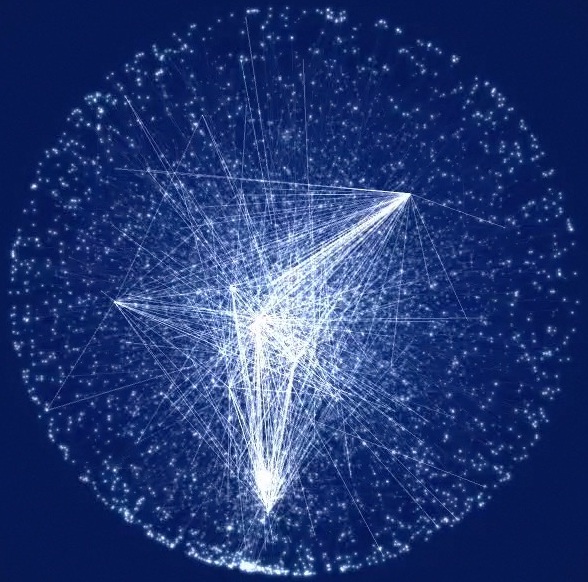
The traditional approach of iterative development, experimental testing, clinical validation, and approval of new drugs are not feasible.

A more realistic strategy relies on drug repurposing, requiring us to identify clinically approved drugs that have a therapeutic effect in COVID-19 patients.

Network medicine has developed and validated a series of computational tools to identify drug repurposing opportunities. Here, we deploy these tools to identify potential drug repurposing candidates for COVID-19.



Network Medicine



Network Medicine

Disease Module Discovery: Identifying the disease modules of multiple phenotypes, pathway analysis, disease gene identification, bioinformatics validation of the modules.

Key Papers: Menche, *Science* (2015). Ghiassian, *PLOS Comp. Biology* (2015).

Drug Target Identification and Repurposing: Identify and validate candidates for drug repurposing.

Key Papers: Guney, *Nature Comm.* (2016) Chen, *Nature Comm.* (2018).

Drug Combinations: Identify drug combinations with higher efficacy than single drugs.

Key Paper: Chen et al, *Nature Comm.* (2019).

Personalized Network Medicine: Placing individual patient data in the context of the disease module, disease heterogeneity, patient classification.

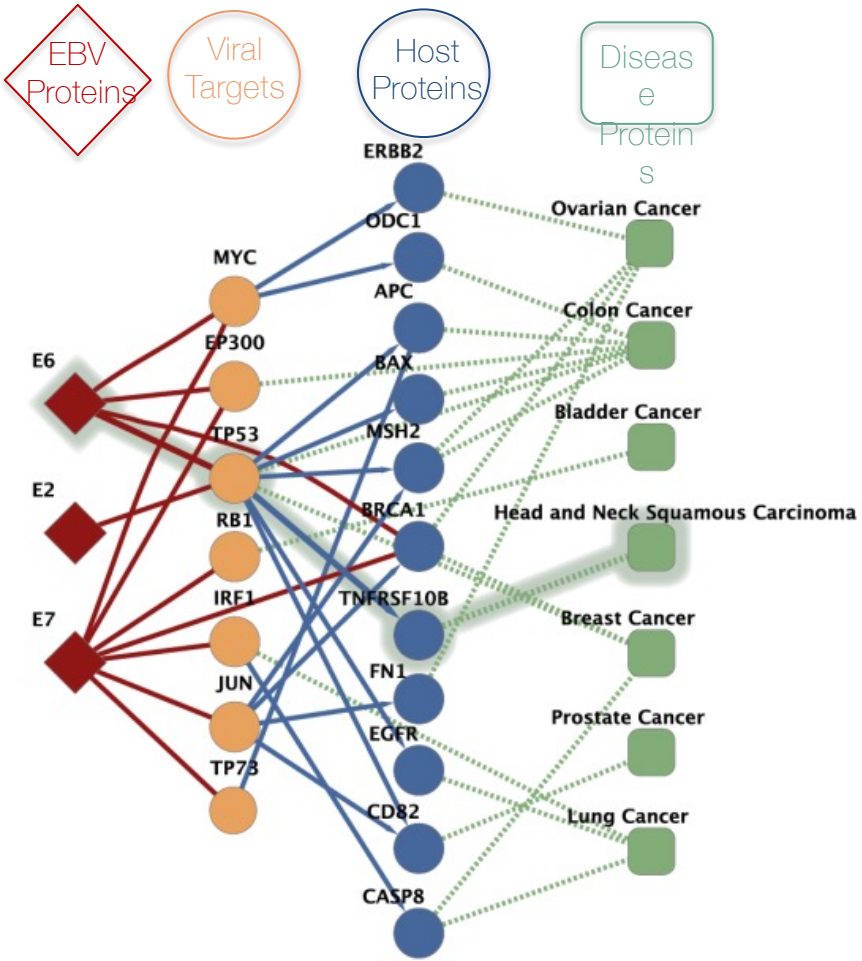
Key Paper: Menche, *Syst. Biol. Appl*. (2017)

Controlling Biological Networks: From subcellular networks to the brain.

Key Papers: Liu, *Nature* (2011); Lee, *Science* (2017); Yan, *Nature* (2017); Towlson, Proc. Roy. Soc. (2018)

Virus-Host Interactions

(EBV and HPV)

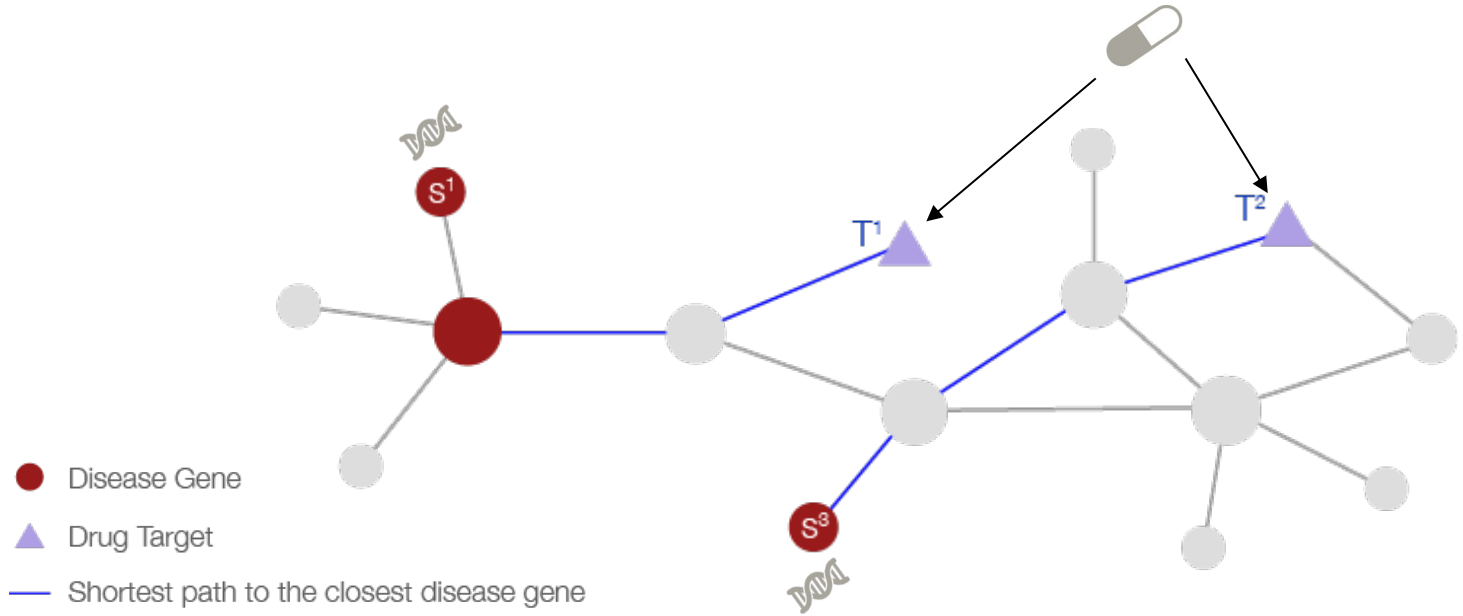


viral protein→ viral targets → disease gene

By inspecting the network neighborhood of the viral targets, we were able to identify the molecular processes disrupted by the virus, and the disease symptoms.

Gulbahce, PLoS Comput Biol. 2012.
Rozenblatt-Rosen, Nature. 2012.

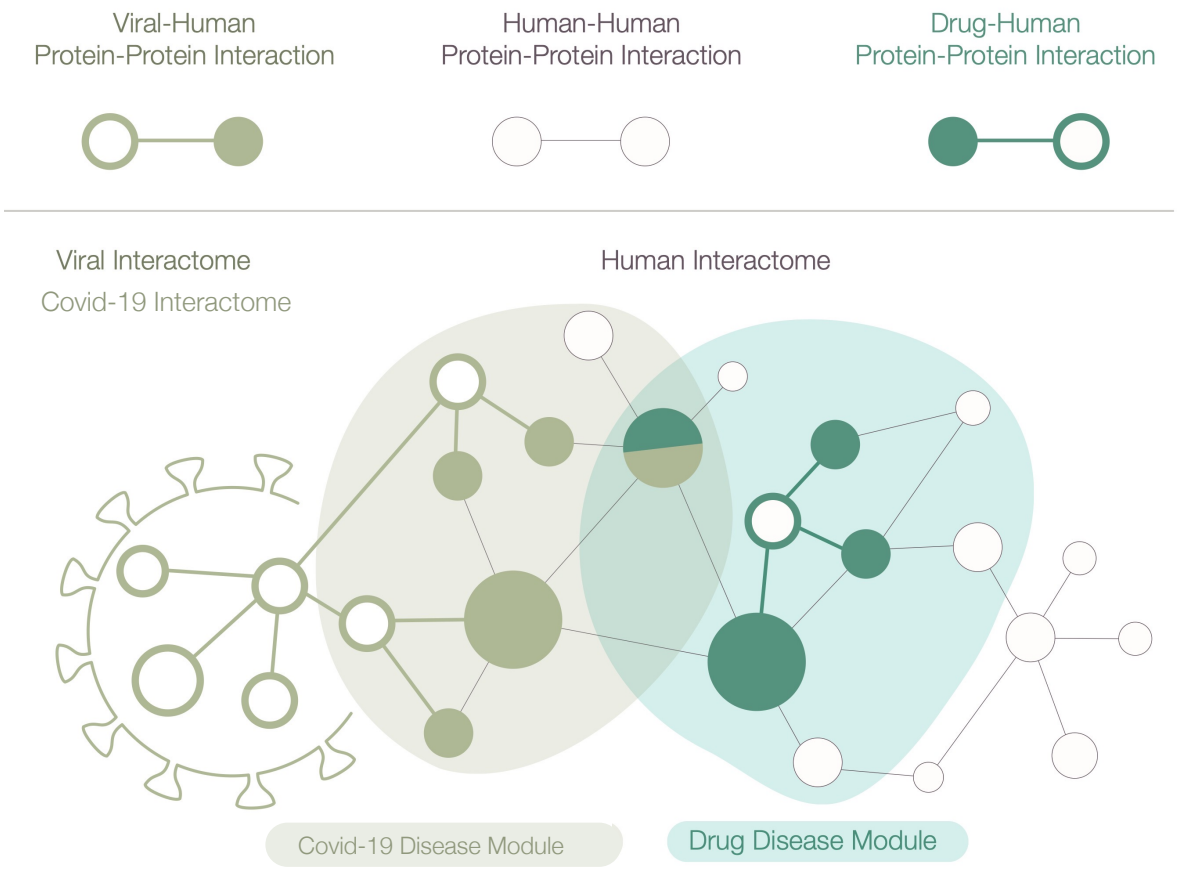
Drug Repurposing: The Proximity Hypothesis



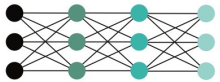
Drugs with targets in the network vicinity of a disease module helps are drug repurposing candidates.

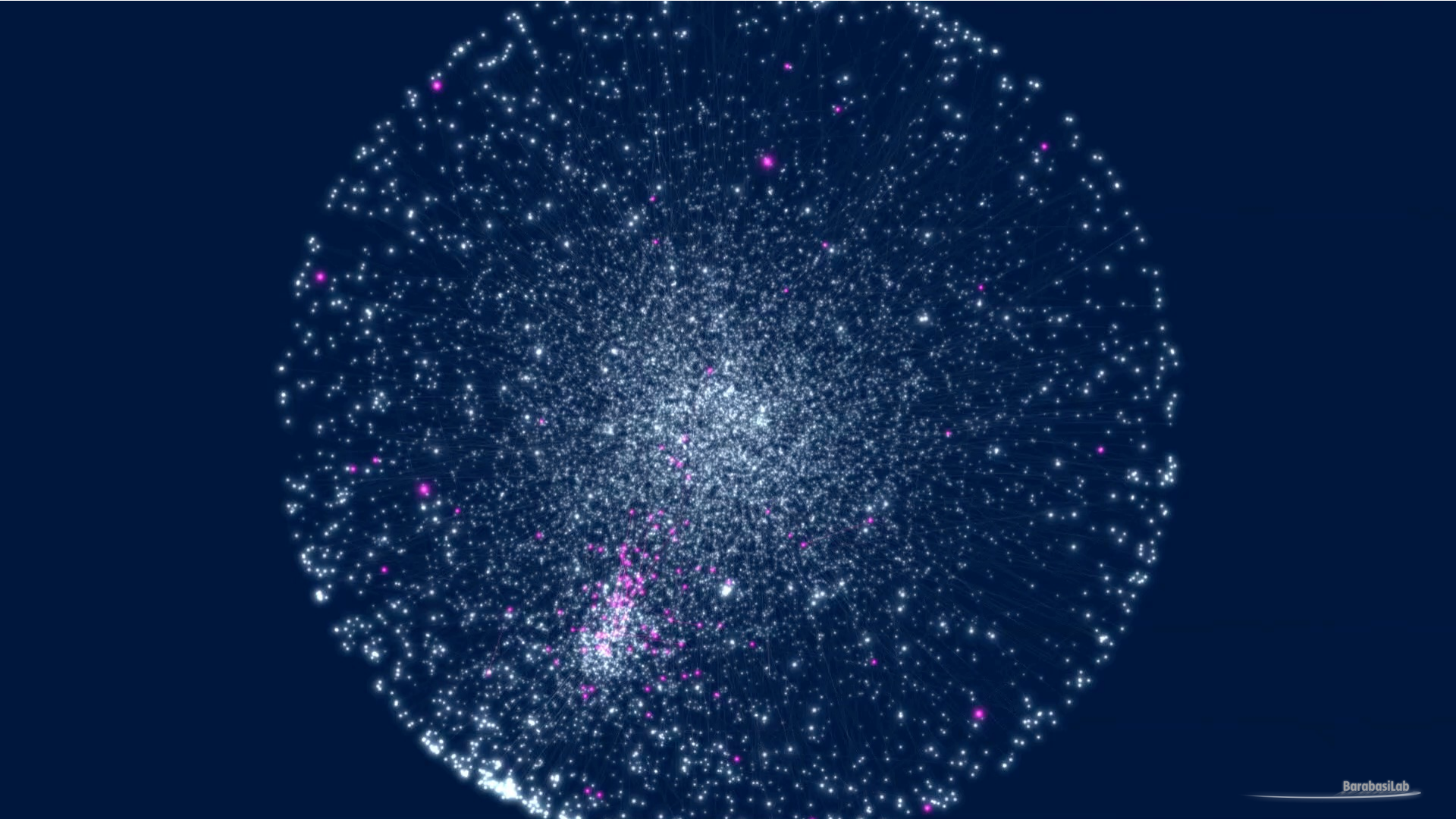
Methodology: Guney et al., Nature Comm.(2016).
Testing using Patient Data: Cheng et al., Nature Comm. (2018)

COVID-19 Disease Module

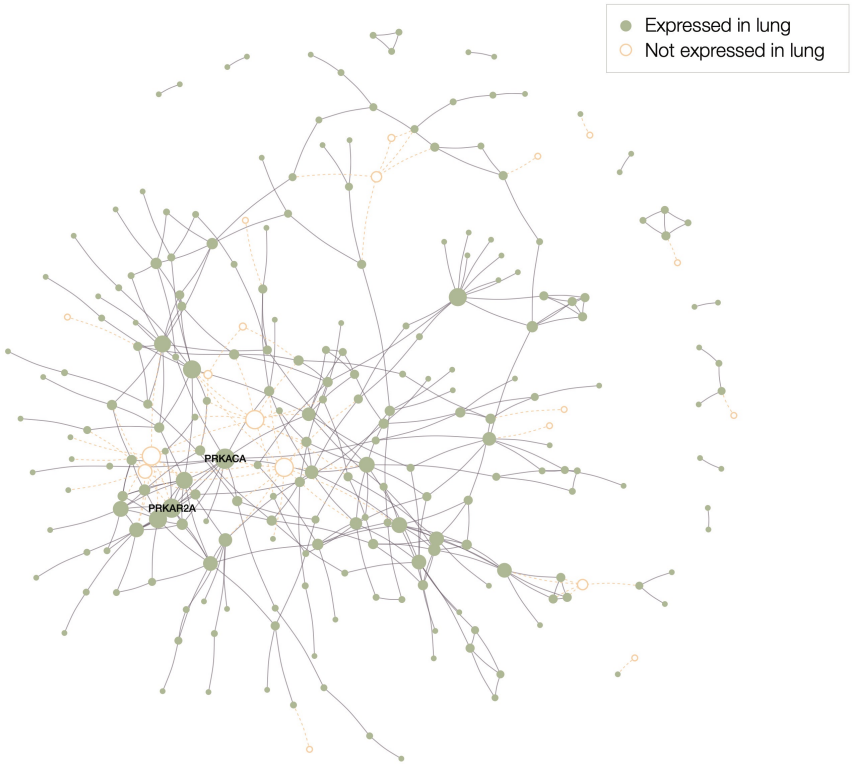


Outline

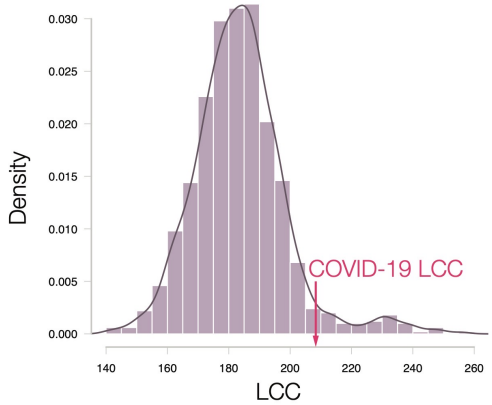
Input Data	Methods	Outcomes
Human Interactome N = 18,508 proteins L = 332,749 PPIs	 Network Proximity 3 pipelines	Infected Tissues/Organs
SARS-COV2 targets 320 human proteins Gordon et al, 2020	 Network Diffusion 5 pipelines	Comorbidity
Drug Targets 7,591drugs 4,187 drug targets DrugBank	 AI Prioritization 4 pipelines	Drug Repurposing & Validation


BarabasiLab

COVID-19 Disease Module



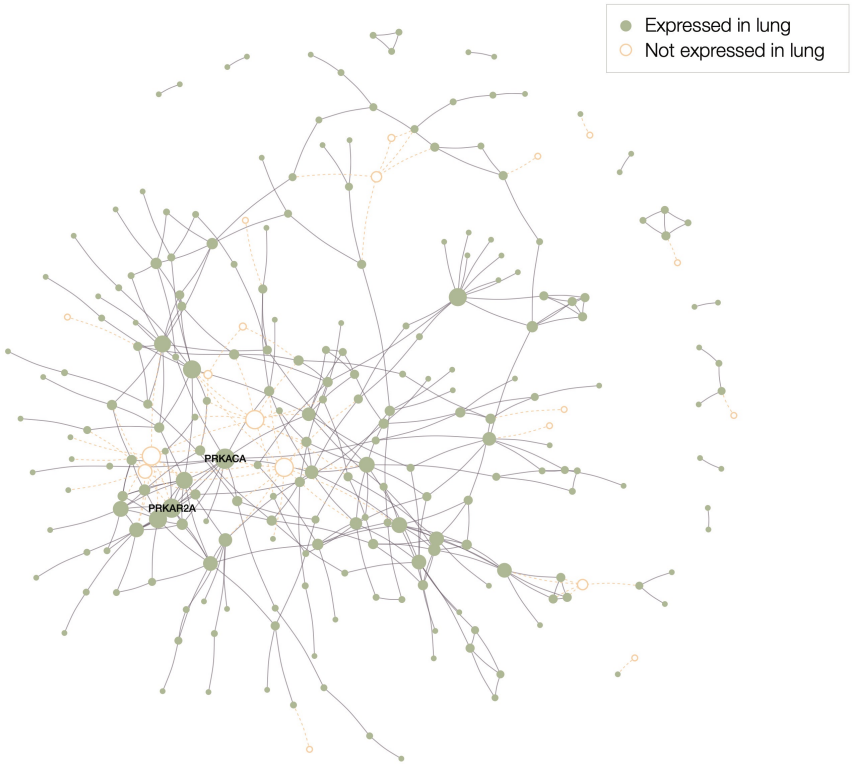
Full Interactome

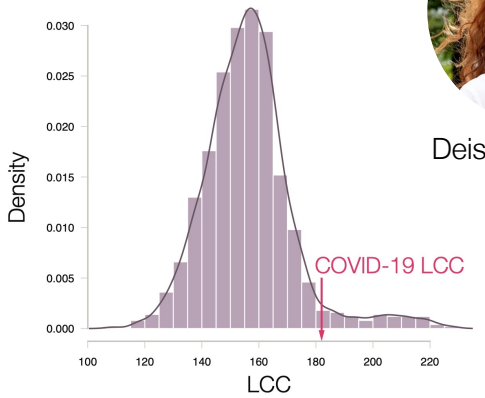


- 332 human proteins to which 26 SARS-CoV2 proteins bind.
- 208 viral targets form a large connected component (LCC)
- Z-Score= 1.65: SARS-CoV2 targets aggregate in the same network vicinity—COVID-19 disease module
- Repurposing candidates must target proteins in the network vicinity of the disease module.

Lonsdale, Nature Genetics. 2013.
Gordon, BioRxiv. 2020.

Tissue Specificity





Deisy Morselli Gysi

- For a disease to be manifest in a tissue, a statistically significant disease LCC must be expressed.
- With GTEx > 5, only 10,823 (58%) proteins in the interactome are expressed in lung.
- The lung specific LCC has a Z-Score= 1.78 (larger than the Z-Score= 1.65 of the LCC in the full-network).

Kitsak, Nature Sci. Rep. 2016; Lonsdale, Nature Genetics. 2013. Gordon, BioRxiv. 2020.

30 tissues express the COVID-19 LCC

Tissue	LCC	Z-Score
Lung	182	1.780
Brain-Hippocampus	149	1.884
Brain-Frontal Cortex	162	1.923
Brain-Cortex	161	1.889
Brain-Hypothalamus	157	1.757
Brain-Spinal cord	169	1.713
Brain-Anterior cingulate cortex	152	1.690
Adrenal Gland	168	1.816
Prostate	183	1.715
Cervix-Endocervix	185	1.801
Ovary	182	1.726
Testis	189	1.794
Uterus	184	1.808
Cervix-Ectocervix	184	1.730
Vagina	185	2.062
Colon-Sigmoid	179	1.870
Colon	179	1.760
Bladder	179	1.799
Esophagus-Mucosa	175	1.757
Pancreas	133	1.908
Artery	178	1.777
Heart-Atrial Appendage	153	1.716
Heart-Left Ventricle	129	1.897
Immortalized cell line	171	2.114
Spleen	173	1.761
Fibroblasts	183	1.843
Skin	178	1.720
Kidney-Cortex	151	1.848
Kidney	167	1.704
All	**208**	**1.658**

Lung
COVID-19 modules is expressed in the respiratory system

multiple brain regions
neurological manifestations, like loss of smell, taste, headache, dizziness, seizure, and skeletal muscular injury

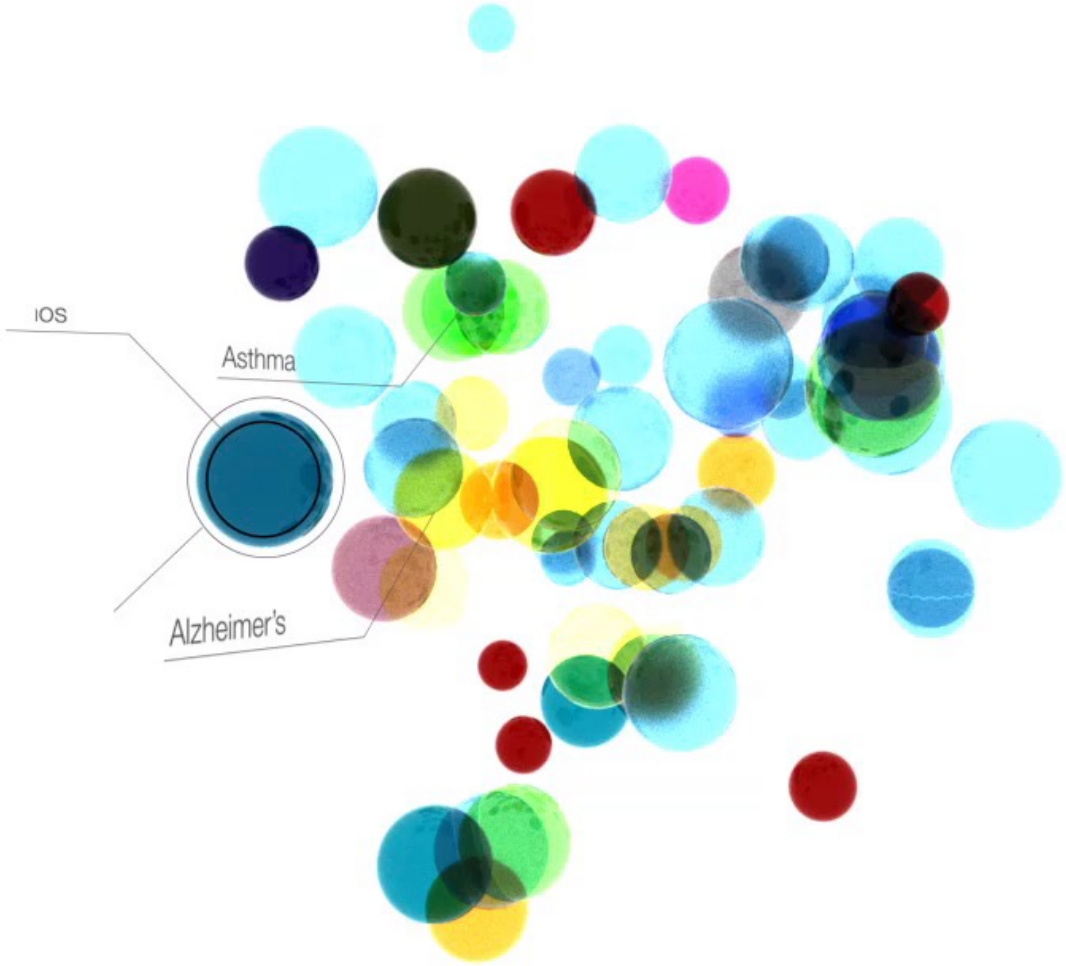
multiple reproductive system tissues

digestive system
consistent with clinical observations such as diarrhea, vomiting and abdominal pain

cardiovascular tissues
infected patients often present significant cardiovascular involvement, and patients with underlying cardiovascular diseases show increased risk of death.

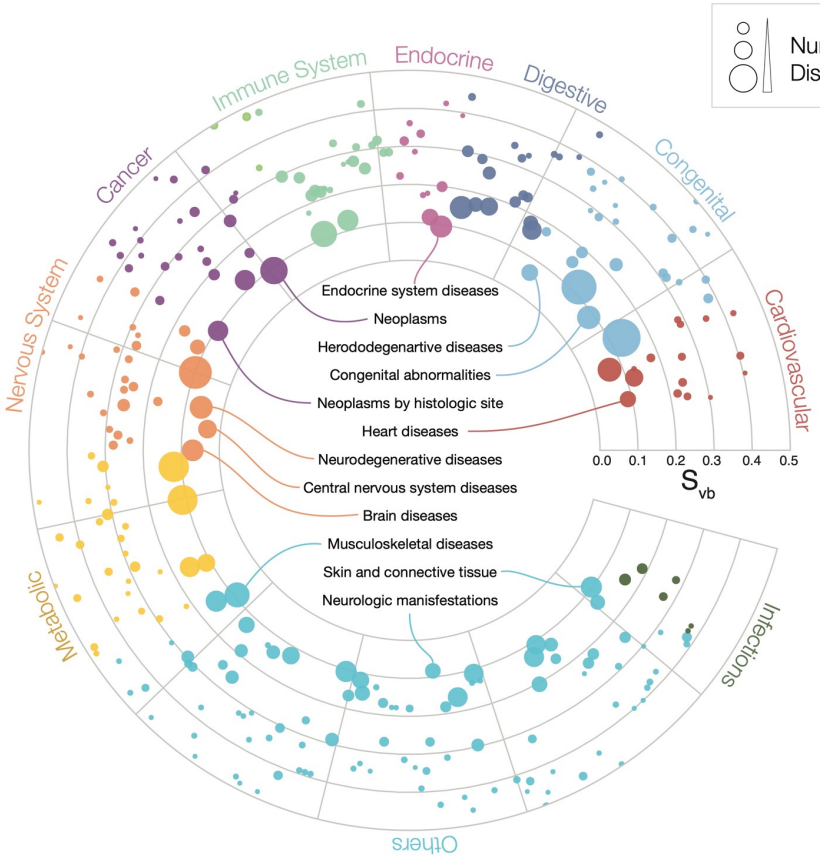
Deisy Morselli Gysi

Xu, Lancet resp med, 2020
Eliezer, JAMA, 2020
Song medRxiv, 2020
Mao, JAMA, 2020
Gu, Gastroenterology. 2020



Mencho et al. Science, 201

Comorbidity



Ítalo Do Valle

The SARS-CoV2 disease module does not directly overlap with any major disease module.

The diseases closest to the COVID-19 proteins include several cardiovascular diseases and cancer, whose comorbidity in COVID-19 patients is well documented.

Neurological diseases, in line with the earlier finding that the virus could be expressed in brain.

Gulbahce, PLoS Comput Biol. 2012.
Menche, Science. 2015.

Methods

A COVID-19 treatment can not be derived from the arsenal of therapies approved for specific diseases

Repurposing strategies focus on drugs previously approved for other pathogens, or on drugs that target the human proteins to which viral proteins bind.

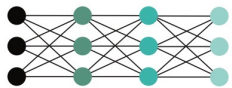
Most approved drugs do not target directly disease proteins but bind to proteins in their network vicinity
[Yildirim, Nature Biotech. 2007]

Identify drug candidates that have the potential to perturb the network vicinity of the COVID-19 disease module.

Implement 3 Network Repurposing Methods.

Network Proximity
3 pipelines

Network Diffusion
5 pipelines



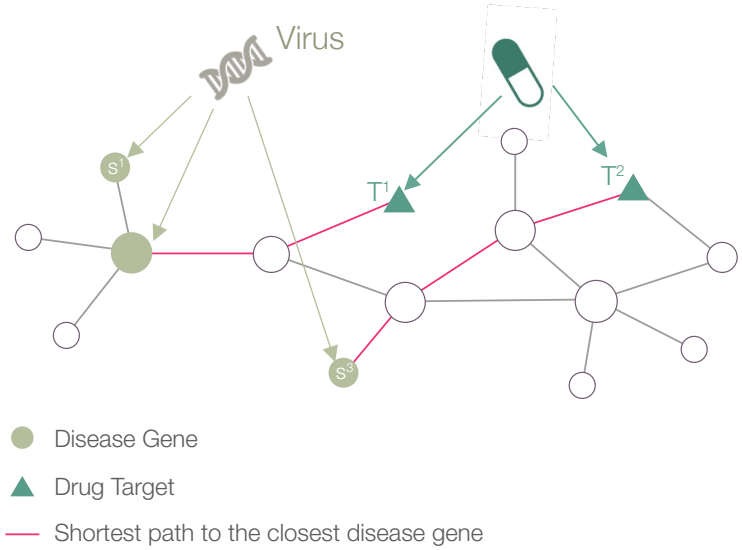
AI Prioritization
4 pipelines

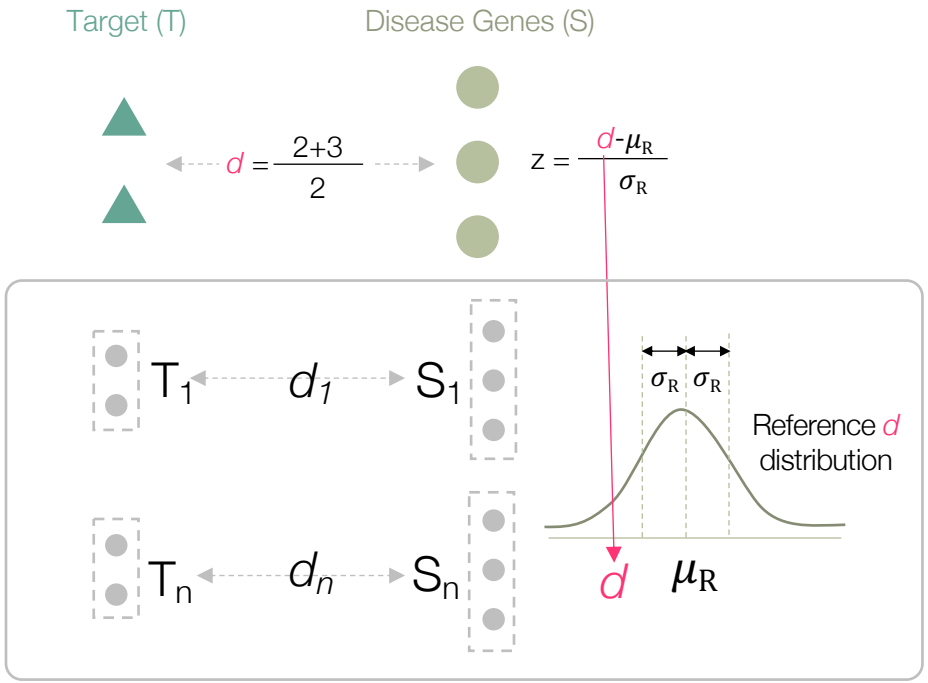
Guney, Nature Comm, 2016.
Lin, IEEE inf. in trans. info. the. 1991
Zitnik, Bioinformatics. 2018
Zitnik, Info. Fision. 2019

Proximity-Based Ranking

Drugs with targets close to disease genes in PPI tend to be efficacious
Rank drugs by relative proximity z-score

Xiao Gan





Random gene sets with same degrees

Guney, Nature Comm. 2016

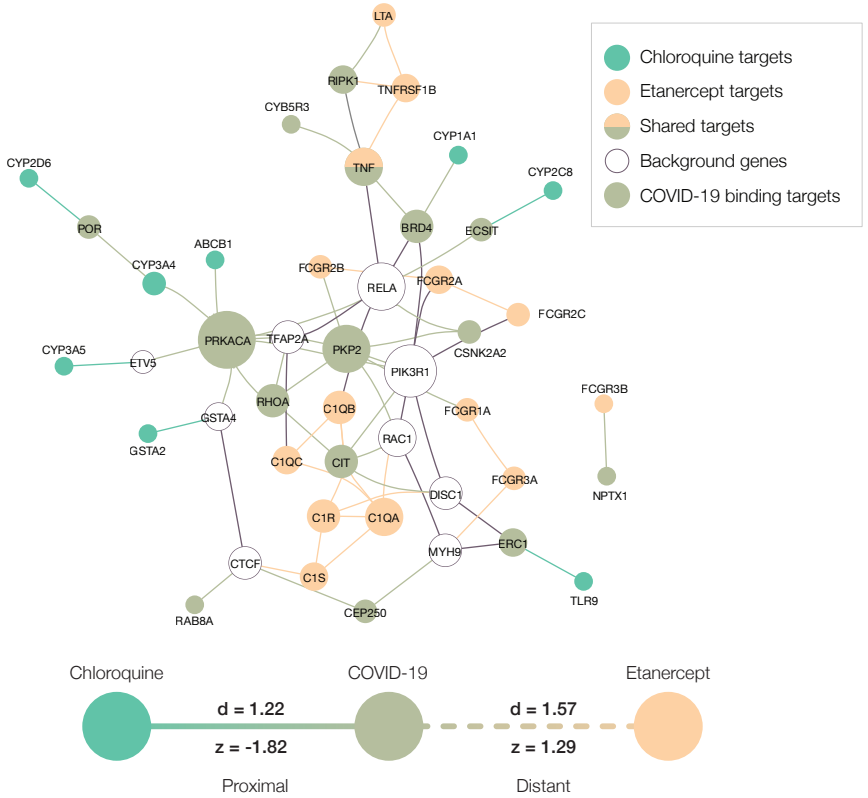
Three Proximity Pipelines

Xiao Gan

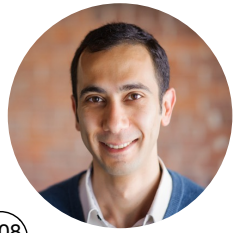
Pipeline P1:
For 6,116 drugs, use all targets to compute proximity z-score.

Pipeline P2:
Disregard targets that are enzymes, carriers or transporters, less related to pharmacological effects (5,550 drugs).

Pipeline P3:
For 793 drugs, compute z-score based on proximity to differentially expressed genes



Diffusion Models: Node Representation


Asher Ameli

Calculate the expected number of times He(A, B) that a random walk starting at node A visits node B.

Represent each node by the vector:

$$He(V_i) = [He(V_i, V_1), He(V_i, V_2), He(V_i, V_3), ..., He(V_i, V_n)]$$

Quantify similarity between a pair of nodes using

a) Manhattan distance:

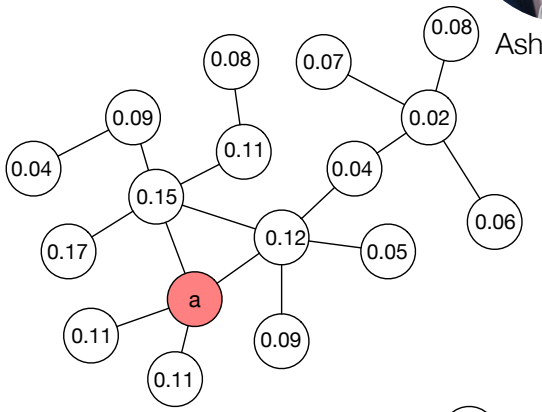
$$DSD(He^A, H^B) = |He^A - He^B|$$

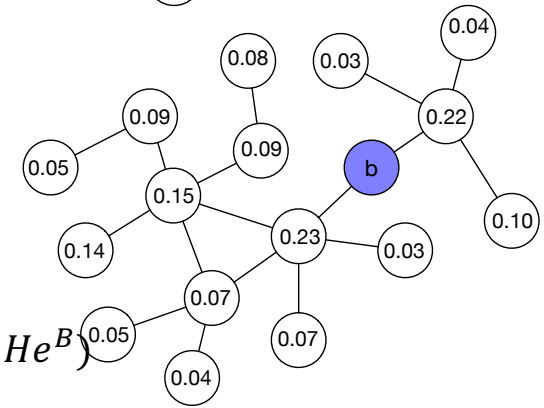
b) Relative entropy (KL divergence):

$$KL(He^A, H^B) = \sum_{x \in V} He^A(x) log \frac{He^B(x)}{He^A(x)}$$

c) Symmetrized KL (JS divergence):

$$JS\,(He^A, H^B) = \frac{1}{2} KL(He^A, M) + \frac{1}{2} KL(He^B, M) \quad , \mathrm{M} = \frac{1}{2}(He^A + He^B)$$





Diffusion Models: Ranking Drugs

Asher Ameli

Five new metrics to calculate the impact of drug targets t on COVID19 targets:

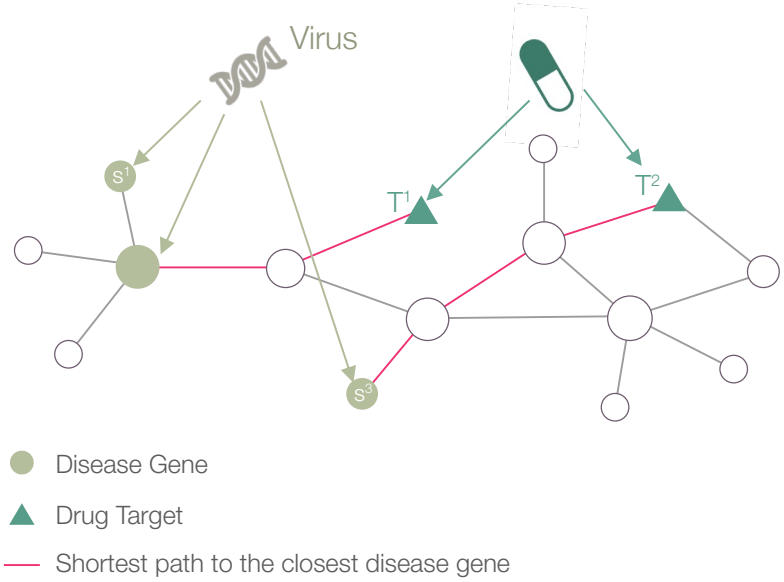
Pipeline D1: $I_{DSD}^{min} = \frac{1}{V}\sum_{t \in T} \min_{v \in V} DSD(t,v)$

Pipeline D2: $I_{KL}^{min} = \frac{1}{V}\sum_{t \in T} \min_{v \in V} KL(t,v)$

Pipeline D3: $I_{KL}^{med} = \frac{1}{V}\sum_{t \in T} \underset{v \in V}{med}\, KL(t,v)$

Pipeline D4: $I_{JS}^{min} = \frac{1}{V}\sum_{t \in T} \min_{v \in V} JS(t,v)$

Pipeline D5: $I_{JS}^{med} = \frac{1}{V}\sum_{t \in T} \underset{v \in V}{med}\, JS(t,v)$

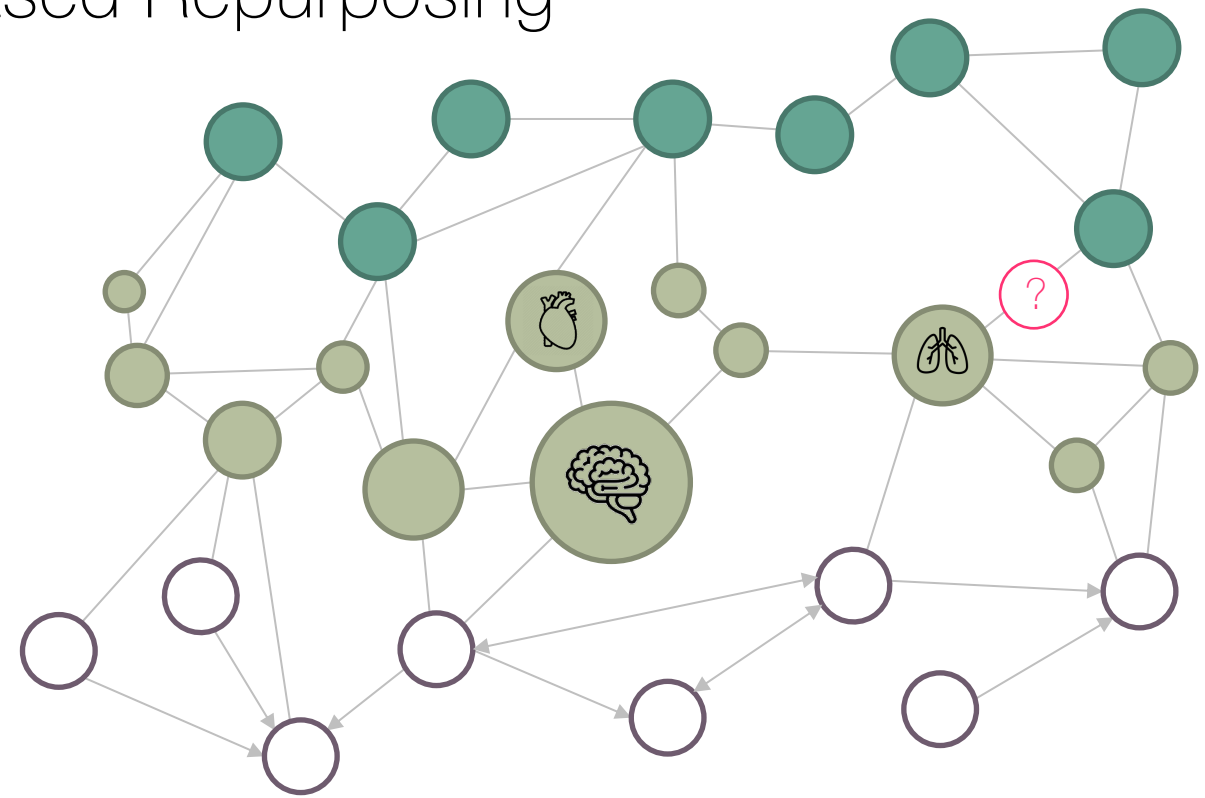


DSD, KL and JS are used to quantify how likely target may impact COVID19 proteins

Lin, IEEE inf in trans. info. the. 1991

AI-Based Repurposing

Marinka Zitnik



Drugs

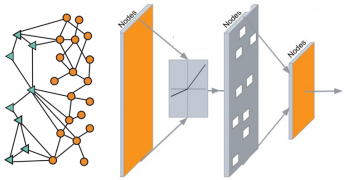
Task
Predict what diseases a new molecule might treat

Virus Proteins

Human Proteins

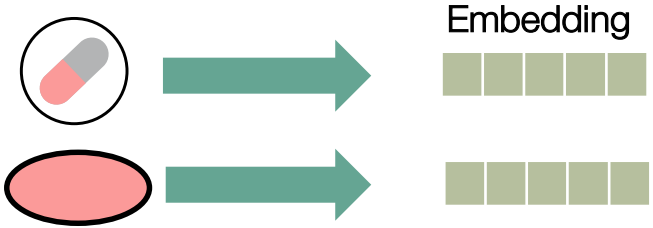
We construct a knowledge graph of biomedical interactions, including drug-target, protein-protein, drug-disease, and disease-protein associations

AI-Net: Graph Neural Network

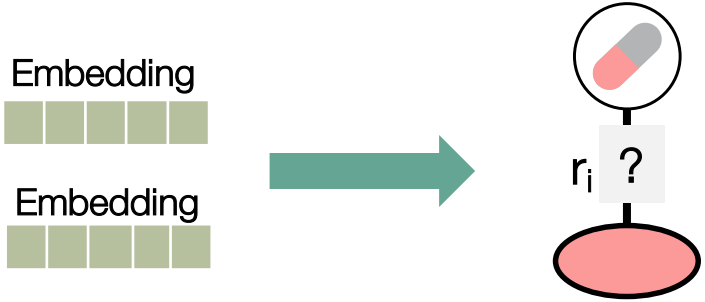


Marinka Zitnik

1. Graph Convolutions: Take a knowledge graph and learn an embedding for every node in the graph



2. Link Prediction: Take the learned embeddings and predict what diseases a given drug might treat



Zitnik, Bioinformatics. 2018
Zitnik, Info. Fision. 2019

AI-Net: Four Prediction Pipelines

Marinka Zitnik

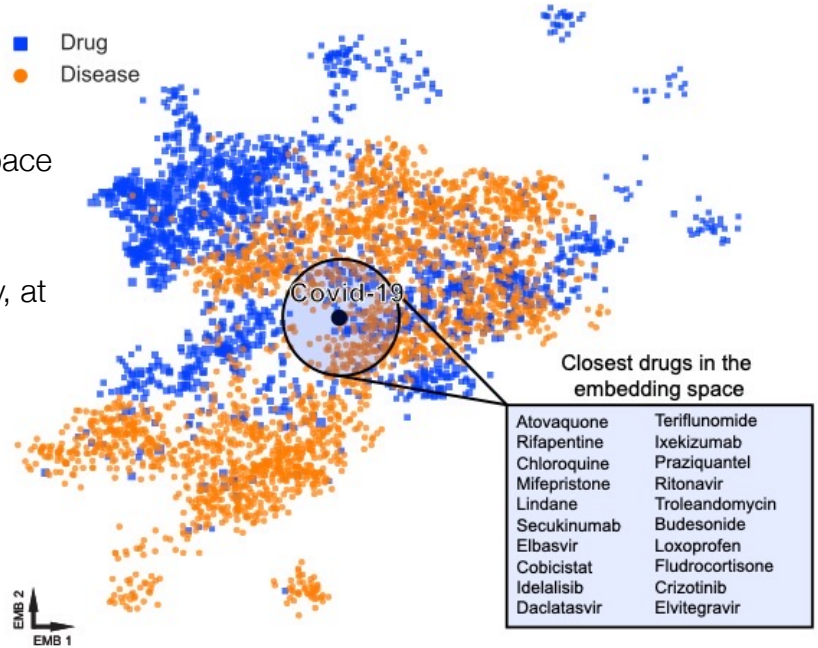
We use **four decoders** to predict disease treatments, *i.e.*, to decode drug-disease links, based on the learned embedding space

Pipeline A1
Search for drugs in the vicinity of COVID-19 by calculating the cosine distance between COVID-19 and all drugs in the decoded embedding space

Pipeline A2
Prevent nodes in the embedding space from packing together too closely, at the loss of the more detailed structure.

Pipeline A3
Force the decoding to concentrate on the very local structure (to the detriment of the overall goal of the exercise).

Pipeline A4
Force the decoding to preserve the broad structure of the embedding space.



Deisy Morselli Gysi

Thee predictive methodologies, offering twelve predictive rankings.

The rankings are not expected to be independent: Start from the same drug and drug-target list and operate on the same interactome.

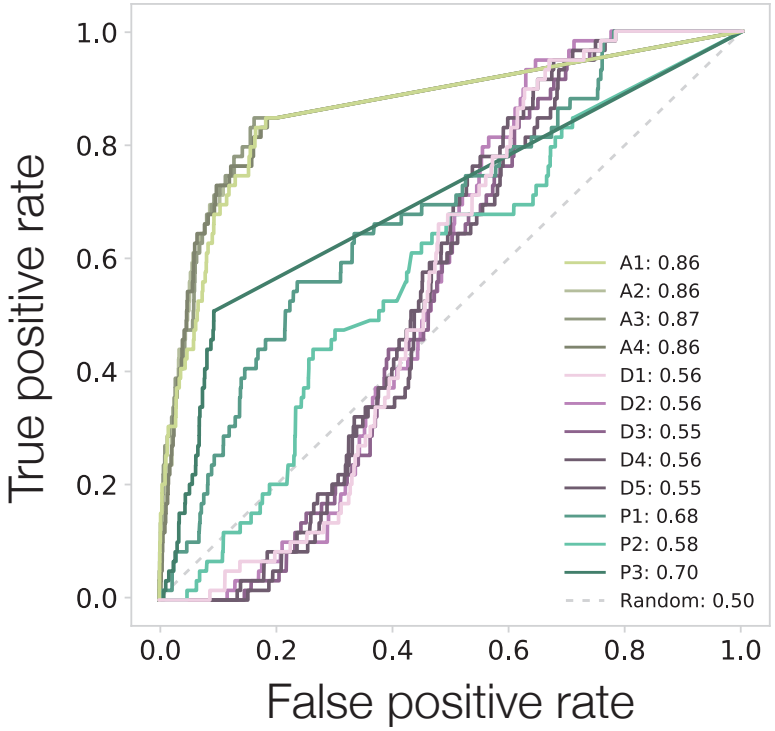
Kendall τ rank correlation of the rankings provided by each pipeline.

Proximity-based pipelines, P1 and P2, show high correlation between each other, as do the AI-Net pipelines (A1-A4), and the diffusion-based pipelines (D1-D5).

Correlations across the three methods are lower, and P3, relying on gene expression, is also uncorrelated with other pipelines.

Different methods offer complementary ranking information

Individual ROC



Onur Varol

We test each pipeline's ability to recover drugs currently in clinical trials for COVID-19 (67 drugs from ClinicalTrials.gov).

The best individual ROC curves are obtained by the AI-based methods.

The second-best performance is provided by the proximity P3. Close behind is P1 with AUC = 0.68

Eliminating some drug targets decreases the AUC to 0.58 (P2)

Diffusion methods offer ROC between 0.55-0.56.

Marinka Zitnik

Combined ROC

Onur Varol

Rank Aggregation Algorithm: Maximize the number of pairwise agreements between the final ranking and each input ranking.

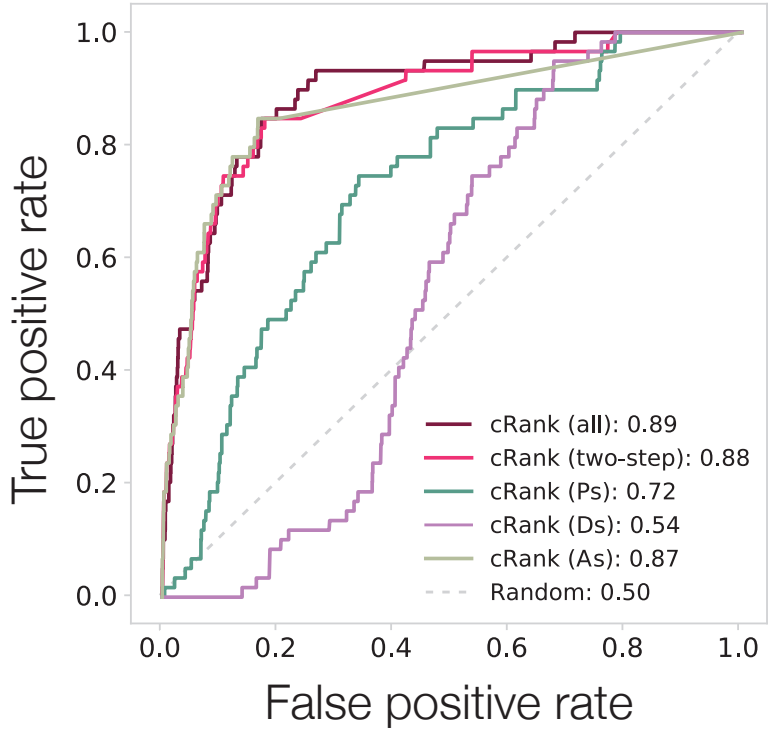
The combined performance of the AI methods is 0.87, the same as A3.

Improvement for proximity pipelines: 0.70 → 0.72.

Combined diffusion pipelines have lower performance (0.54 vs 0.56, for D1, D2, and D4).

Combining all 12 pipelines, we obtain AU=0.89, the highest of any individual or combination-based pipelines,

Individual pipelines offer complementary information harnessed by the combined ranking.



Zitnik, Nat Comm. 2018

86 drugs selected from the top 10% of the rank list.

Respiratory drugs (e.g., theophylline, montelukast).

Cardiovascular systems (e.g., verapamil, atorvastatin).

Antibiotics used to treat viral (e.g., ribavirin, lopinavir), parasitic (e.g., hydroxychloroquine, ivermectin, praziquantel), bacterial (e.g., rifaximin, sulfanilamide), mycotic (e.g., fluconazole), and mycobacterial (e.g., isoniazid) infections.

Immunomodulating/anti-inflammatory drugs (e.g., interferon-β, auranofin, montelukast, colchicine)

Anti-proteasomal drugs (e.g., bortezomib, carfilzomib)

Less obvious choices: aminoglutethimide, melatonin, levothyroxine, calcitriol, selegiline, deferoxamine, mitoxantrone, metformin, nintedanib, cinacalcet, and sildenafil.

◯ # of Clinical trials from ClinicalTrials.gov

Joseph Loscalzo

	Drug	C-rank
(20)	Ritonavir	1
	Isoniazid	2
	Troleandomycin	3
	Cilostazol	4
(76)	Chloroquine	5
	Rifabutin	6
	Flutamide	7
(2)	Dexamethasone	8
	Rifaximin	9
	Azelastine	10
	Folic Acid	16
	Rabeprazole	27
	Methotrexate	32
	Digoxin	33
	Theophylline	34
	Fluconazole	41
	Aminoglutethimide	42
(67)	Hydroxychloroquine	44
	Methimazole	47
(1)	Ribavirin	49
(1)	Omeprazole	50
	Bortezomib	53
	Leflunomide	54
	Dimethylfumarate	55
(4)	Colchicine	57
	Quercetin	63
	Mebendazole	67

	Drug	C-rank
	Mesalazine	69
	Pentamidine	92
	Verapamil	98
	Melatonin	109
	Griseofulvin	112
	Auranofin	118
(1)	Atovaquone	124
	Montelukast	131
	Romidepsin	138
(1)	Cobicistat	141
(17)	Lopinavir	146
	Pomalidomide	155
	Sulfinpyrazone	157
(1)	Levamisole	161
	Calcitriol	164
(1)	Interferon-β-1a	173
	Praziquantel	176
(1)	Ascorbic acid	195
	Fluvastatin	199
(1)	Interferon-β-1b	203
	Selegiline	206
(1)	Deferoxamine	227
	Ivermectin	235
(1)	Atorvastatin	243
	Mitoxantrone	250
	Glyburide	259
(2)	Thalidomide	262

	Drug	C-rank
	Sulfanilamide	26[illegible]
	Hydralazine	269
	Gemfibrozil	281
(4)	Ruxolitinib	284
	Propranolol	297
	Carbamazepine	301
	Doxorubicin	309
	Levothyroxine	329
	Dactinomycin	335
	Tenofivir	338
	Tadalafil	339
	Doxazosin	367
	Rosiglitazone	397
	Aminolevulinic acid	398
	Nitroglycerin	418
	Metformin	457
(1)	Nintedanib	466
	Allopurinol	471
	Ponatinib	491
(1)	Sildenafil	493
	Dapagliflozin	504
	Nitroprusside	515
	Cinacalcet	553
	Mexiletine	559
	Sitagliptin	706
	Carfilzomib	765
(1)	Azithromycin	786

Validation Case Studies: Connectivity Map

Ítalo Do Valle

Measured the overlap between perturbed genes and COVID-19 targets for 59 of the 81 repurposing candidates in the Connectivity Map.

Mitoxantrone (antineoplastic): 75 (22%) of the COVID-19 targets have a significant overlap with the 2,440 genes perturbed by the drug in the lung (see Figure).

Statistically significant overlap for 43 drugs (random: 13 ± 7): repurposing candidates effectively perturb the COVID-19 module.

Highest number of perturbed COVID-19 targets: carfilzomib (162), flutamide (162), and bortezomib (162).

For lung the drugs with the highest overlap with COVID-19 targets are mitoxantrone and ponatinib.

Subramanian, Cell. 2017.
Lamb, Science. 2006.

Validation: Suppressing COVID-19 Expression

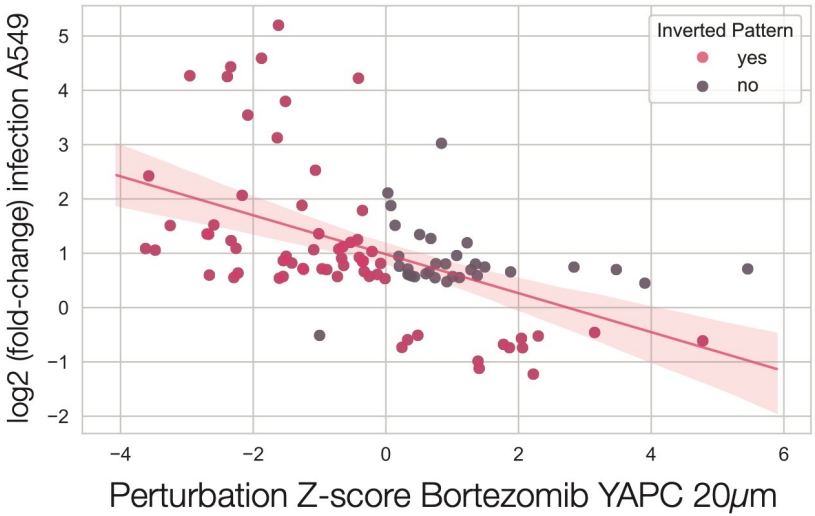
Ítalo Do Valle

Counteract the gene expression perturbations caused by the virus: Down-regulate genes up-regulated by the virus or vice versa?

120 differentially expressed genes (DEGs) in the SARS-CoV2 infected of the A549 cell line.

Bortezomib treatment of the cell line YAPC (20 μM) counteracts the effects of the SARS- CoV2 infection for 65 genes (see Figure, Spearman correlation ρ = −0.58).

22 drugs in the Connectivity Map have ρ < 0, indicating that they counteract the effects of the infection (random selections)



Subramanian, Cell. 2017.
Lamb, Science. 2006.

Discussion

We ranked existing drugs based on their expected efficacy for COVID-19 patients. This does not mean that drugs that did not make our final list could not have efficacy, or that they must be excluded from further consideration.

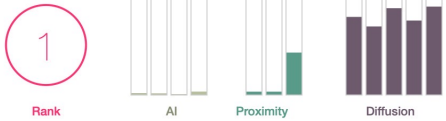
As the input data improves, so will our ranking, and we may develop a case for other drugs, currently not listed:

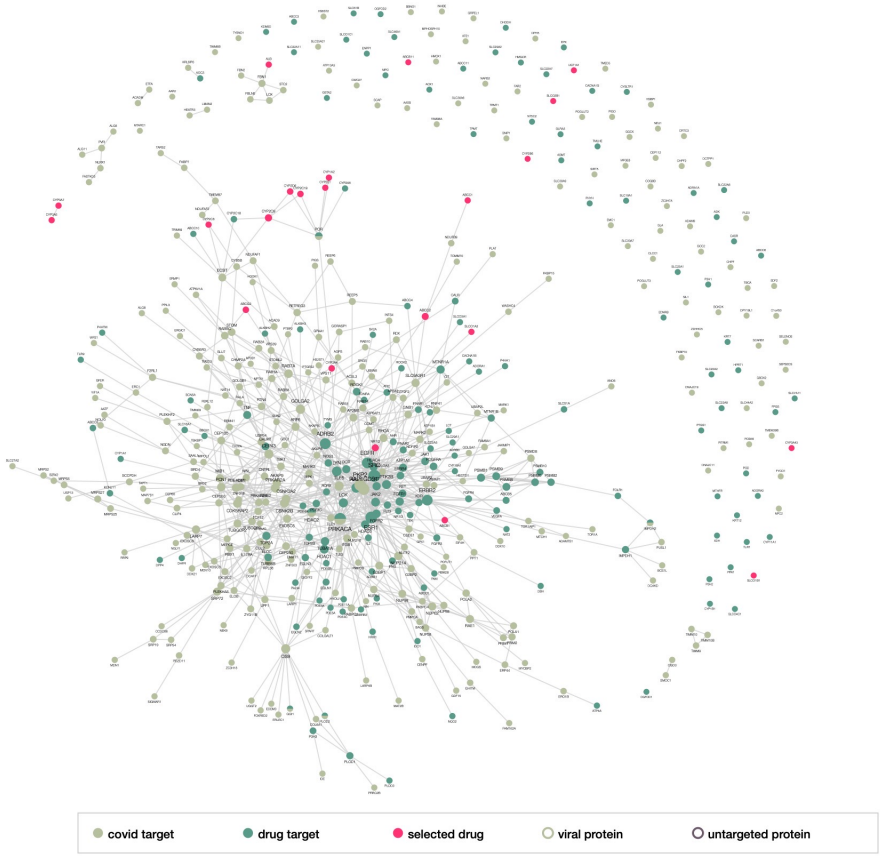
- New virus-host binding interactions (experimental, or AI predicted)
- Microarray data of COVID-19 patients
- New drugs added to DrugBank
- Improved drug target identification

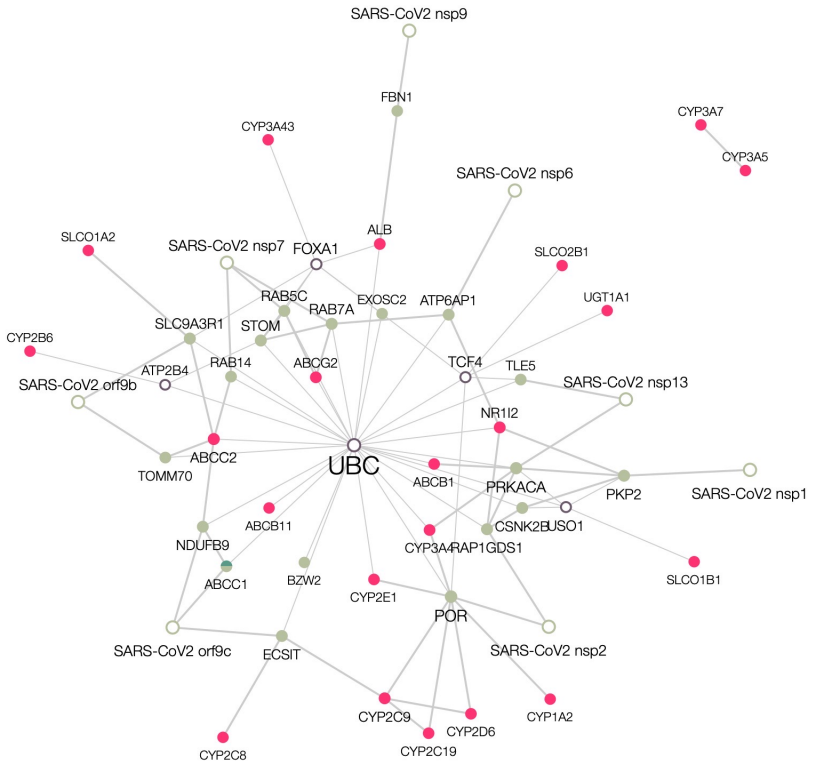
The proposed methodology is general, allowing us to profile the potential efficacy of any drug or a family of drugs, whether they are included in our current reference list.

Ritonavir

20 clinical trials

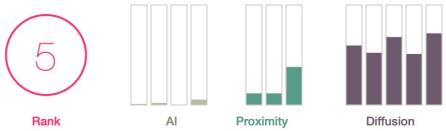


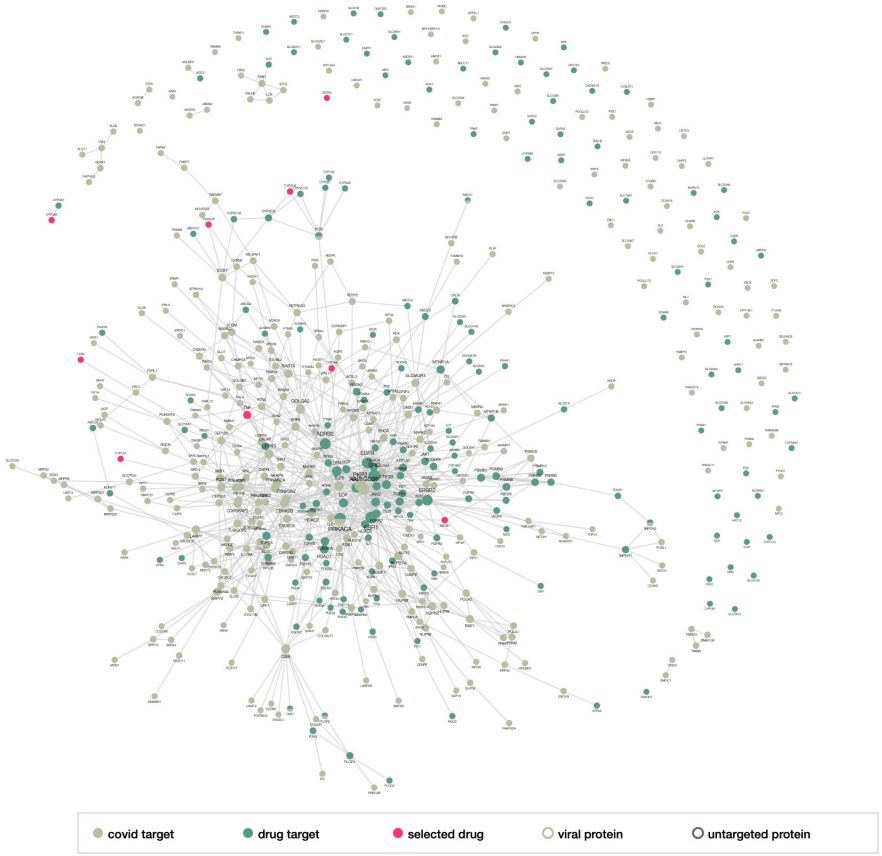


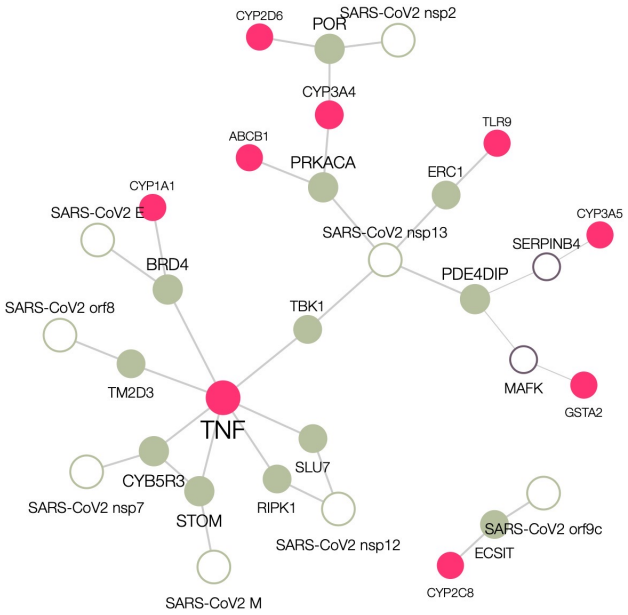


Chloroquine

76 clinical trials

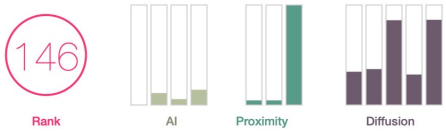


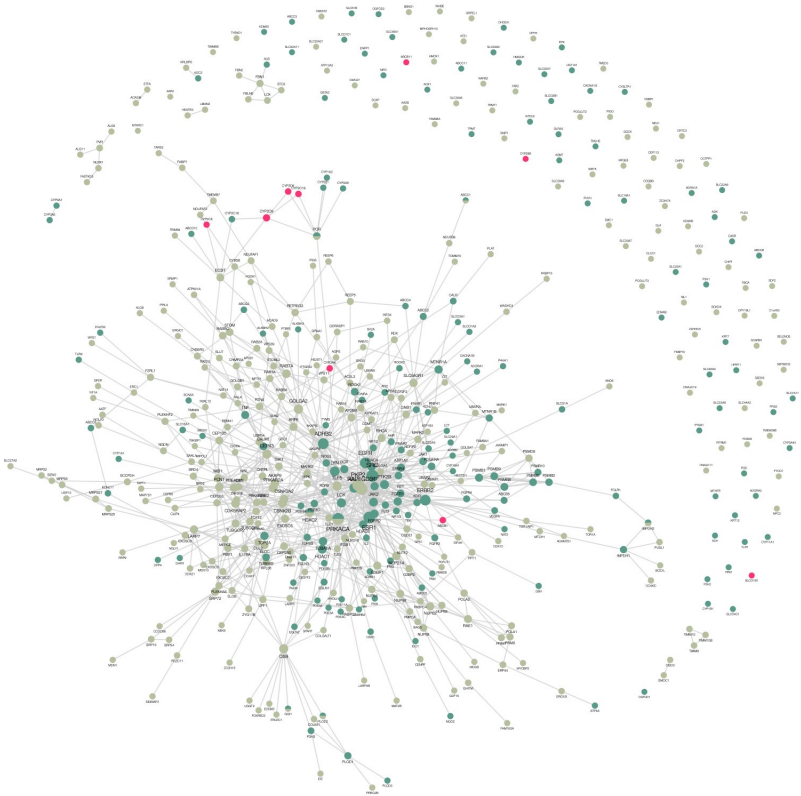


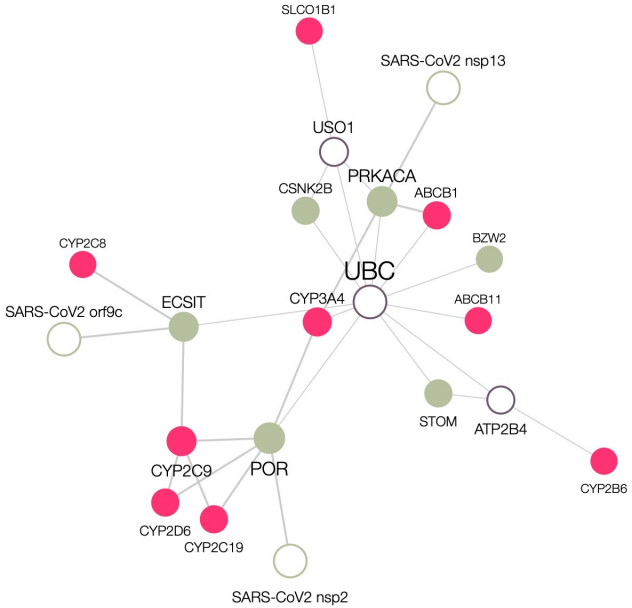


Lopinavir

17 clinical trials

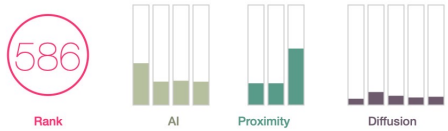


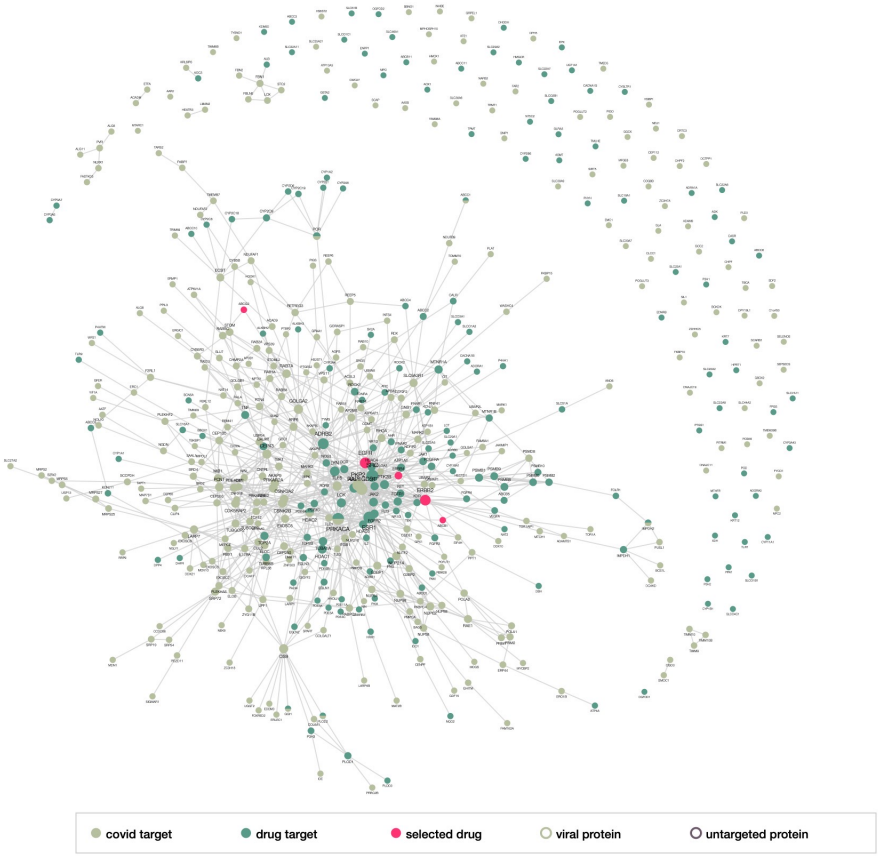


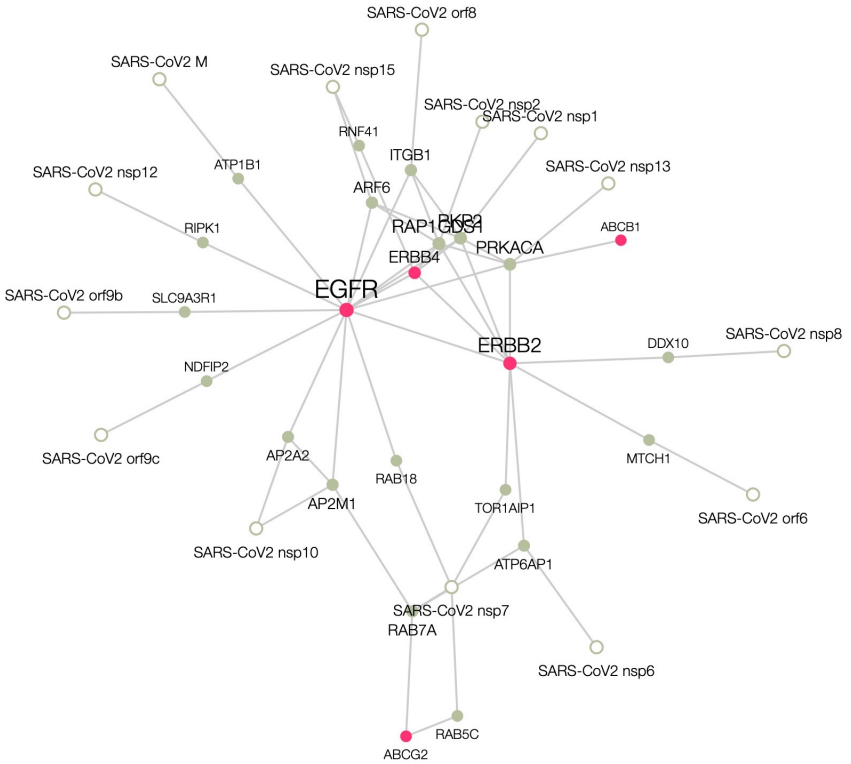


Afatinib

0 clinical trials







Discussion

The predictive pipelines select drugs that are positioned to perturb effectively the COVID-19 disease module.

The perturbations may block the virus' ability to invade the host cells or limit the molecular level disruption caused by the infection. Others, however, may aggravate the symptoms and the seriousness of the phenotype.

Molecular experiments can help test the efficacy of the drugs for COVID-19 infected cell lines.

As these drugs have known side effects and toxicities, it may be possible to move some directly into clinical trials, cognizant of the possibility that these approved drugs may exert unique toxicities in the setting of this novel infection, an outcome that can only be identified in clinical trial.

What is Next?

Waiting for the drugs to arrive, to be screened at the National Emerging Infectious Diseases Laboratories @ Boston University

Working to understand why the pipelines predict what they do – what does the AI see compared to proximity-based methods?

Where does the predictive power come from?

Planning to integrate new data, as it becomes available, and run robustness checks.

Feel free to contact us, if you want to follow up on any of our predictions, and you need further data or details.

Manuscript available at: https://arxiv.org/abs/2004.07229

Other Ongoing Efforts

@KroganLab

KROGAN LAB

Jan Baumbach, TU Munich.

CoVex TUM

Feixiong Cheng, Cleveland Clinic

Cleveland Clinic
Lerner Research Institute

Madhavi Ganapathiraju, U. Pitt.

Wiki-CORONA
Search for PPIs related to Coronavirus Infection:

https://hagrid.dbmi.pitt.edu/corona/

Team

Deisy Morselli Gysi
Northeastern University

Italo Do Valle
Northeastern University

Marinka Zitnik
Harvard University

Asher Ameli
Scipher Medicine

Xiao Gan
Northeastern University

Onur Varol
Northeastern University

Helia Sanchez
Scipher Medicine

Rebecca Baron
Brigham&Women's Hospital
HMS

Dina Ghiassian
Scipher Medicine

Joseph Loscalzo
Harvard Medical School

BarabasiLab.com
barabasi@gmail.com

Acknowledgments

This work was supported, in part, by NIH grants HG007690, HL108630, and HL119145, and by AHA grant D700382 to J.L; A.L.B is supported by NIH grant 1P01HL132825, American Heart Association grant 151708, and ERC grant 810115-DYNASET.

We thank Nicolette Lee and Grecia Morales for providing support, Marc Santolini for suggestions in the diffusion-based methods.

J.L. and A.L.B are co-scientific founder of Scipher Medicine, Inc., which applies network medicine strategies to biomarker development and personalized drug selection. A.L.B is the founder of Nomix Inc. and Foodome, Inc. that apply data science to health; O.V and D.M.G are scientific consultants for Nomix Inc. I.D.V is a scientific consultant for Foodome Inc.