http://druggablegenome.net/ http://datascience.unm.edu/

➔ toprea[at]salud.unm.edu ←

# AN ACADEMIC PERSPECTIVE ON DRUG DISCOVERY & REPOSITIONING FOR COVID-19

Tudor I. Oprea

with contributions from

Sorin Avram, Giovanni Bocci, Cristian Bologa, Steven Bradfute, Lars Juhl Jensen, Praveen Kumar, Jordi Mestres, Douglas J. Perkins, Vishal B. Siramshetty, Anna Waller, Jeremy J. Yang, Gergely Zahoranszky-Kohalmi & many others

Funding: U24 CA224370, U24 TR002278, U01 CA239108 (NIH)

4/23/2020 XXVI Symposium "Bioinformatics and Computer-Aided Drug Discovery" Organized by the V.N. Orekhovich Scientific Research Institute of Biomedical Chemistry --- Updated 4/23/20



## THREE PILLARS OF DRUG DISCOVERY



- There are 3 main pillars at the root of successful drug discovery programs.
- Informatics, Data Science and Machine Learning ("AI" according to the hype cycle) are successfully used, as follows:
- Diseases: Significant improvements in EMR processing, nosology, ontology, and EMR-based ML for Dx & mechanisms
- **Targets:** Knowledge graph methods, coupled with ML, for biological discovery, target selection & validation
- **Drugs:** From virtual screening to vaccine design, therapeutic modalities benefit from predictive methods across the board
- **IDG** is developing methods applicable to each of these 3 areas



Diseases image credit: Julie McMurry, Melissa Haendel (OHSU). All other images credit: Nature Reviews Drug Discovery

2/5/20 revision



January 2017 volume 16 no. 1 www.nature.com/reviews



DRUG TARGETS A comprehensive map of the molecular targets of approved drugs Inflammatory and autoimmune diseases Targeting colony stimulating factors

## A COMPREHENSIVE MAP OF MOLECULAR DRUG TARGETS

We curated 667 human genome-derived proteins and 226 pathogen-derived biomolecules through which 1,578 US FDAapproved drugs act.

This set included 1004 orally formulated drugs as well as 530 injectable drugs (approved through June 2016).

Data captured in DrugCentral (link)



R. Santos et al., Nature Rev. Drug Discov. 2017, 16:19-34 link



May 2018 volume 17 no. 5 www.nature.com/reviews

## ILLUMINATING THE DRUGGABLE GENOME



	GPCRs	U24 DK116195: Bryan Roth, M.D., Ph.D. (UNC) Brian Shoichet, Ph.D. (UCSF)		
RFA-RM-16-026 (DRGC)	Ion Channels	U24 DK116214: Lily Jan, Ph.D. (UCSF) Michael T. McManus, Ph.D. (UCSF)		
	Kinases	U24 DK116204: Gary L. Johnson, Ph.D. (UNC)		
RFA-RM-16-025 (RDOC)	Outreach	U24 TR002278: Stephan C. Schürer, Ph.D. (UMiami) Tudor Oprea, M.D., Ph.D. (UNM) Larry A. Sklar, Ph.D. (UNM)		
RFA-RM-16-024 (KMC)	Data	U24 CA224260: Avi Ma'ayan, Ph.D. (ISMMS) U24 CA224370: Tudor Oprea, M.D., Ph.D. (UNM)		
RFA-RM-18-011 (CEIT)	Tools	U01 CA239106: N Kannan, PhD & KJ Kochut (UGA) U01 CA239108: PN Robinson, MD PhD (JAX), CJ Mungall (LBL), T Oprea (UNM) U01 CA239069: G Wu, PhD (OHSU), PG D'Eustachio PhD (NYU), Lincoln D Stein, PhD (OICR)		



T. Oprea et al., Nature Rev. Drug Discov. 2018, 17:317-332 link

#### TARGET DEVELOPMENT LEVELS



- Most protein classification schemes are based on structural and functional criteria.
- For therapeutic development, it is useful to understand how much and what types of data are available for a given protein, thereby highlighting well-studied and understudied targets.
- Tclin: Proteins annotated as drug targets
- Tchem: Proteins for which *potent* small molecules are known
- Tbio: Proteins for which biology is better understood
- Tdark: These proteins lack antibodies, publications or Gene RIFs

**2020 Update:** Tdark 31.2%; Tbio 57.7%; Tchem 8%; Tclin 3.1%



T. Oprea et al., Nature Rev. Drug Discov. 2018, 17:317-332 link

2/10/20 revision

## IDG KMC ANNOTATION PROCESS



#### **Further information**

Email: idg.rdoc@gmail.com Follow: @DruggableGenome URLs:

https://druggablegenome.net/ https://commonfund.nih.gov/idg/



**IDG Knowledge User-Interface** Email: pharos@mail.nih.gov Follow: @IDG\_Pharos URL: <u>https://pharos.nih.gov/</u>

#### IDG databases are interfaced in UniProt

https://www.uniprot.org/news /2019/09/18/release



#### GTEx, LINCS, IMPC: Data from 3 CommonFund programs is already in Pharos

T. Oprea, Mammalian Genome, 2019, 30:192-200 https://bit.ly/2NUK0BK

4/25/19 revision

#### nature reviews drug discovery



## RARE DISEASES: AN INFORMATICS SURVEY

- We revised the number of RDs from ~7,000 to 10,393 using <u>Disease Ontology</u>, <u>OrphaNet</u>, <u>GARD</u>, <u>NCIT</u>, <u>OMIM</u> and the <u>Monarch</u> <u>Initiative MONDO</u> system
- We also pointed out the lack of a uniform definition for rare diseases, and called for coordinated efforts to precisely define them
- We surveyed therapeutic modalities (<u>link</u>) available to translate advances in the scientific understanding of rare diseases into therapies, and discussed overarching issues in drug development for rare diseases.



Haendel M, et al. Nature Rev. Drug Discov. 2020 19:77-78 link

## TAKE HOME MESSAGE THERE IS A KNOWLEDGE DEFICIT

~31% of the proteins remain understudied (Tdark, ignorome) that number is steadily decreasing

~11.1% of the Proteome (Tclin & Tchem) are currently targeted by small molecule probes and drugs – *that number is slowly increasing* 

With help from rare disease patient advocacy groups, rare disease research is likely to witness a significant increase in translation



#### **COVID-19 CLINICAL TRIALS**

- To date, no drug is expressly approved for SARS-CoV-2 infections except the emergency authorization for (hydroxyl)chloroquine.
- Until such time that effective vaccines and/or therapeutics are approved, our "best guess" is "<u>drug repositioning</u>" (aka drug repurposing) followed by drug discovery
- As of 4/15/20, there were 585 clinical trials for "COVID-19".
- Of these 585 clinical trials, 274 listed a "drug" intervention; of these, 259 reference an actual medicine; 230 are "interventional".
- Examples of non-medicine "drug" entries: BCG vaccine; gargle/mouthwash...



#### CURATING COVID-19 CLINICAL TRIALS

- Manual curation (by intervention), e.g., identify experimental (novel) drug vs. already approved drug; reconcile spelling errorls (e.g., hidroxicloroquin; abidol)
- Five general categories: Placebo (69), Antiviral (54), Experimental (29), Repurposed (201), Biologic (50)
- Twenty-one "specific" categories: HCQ (76/23), CQ (11), Azithromycin (25/23), -navir (18), Oseltamivir (4), Favipiravir (3), Umifenovir (4), Remdesivir (7), -tinib (10), RAS drugs (12), NSAIDs (4), Steroid (17), TMPRSS2 (4), Traditional Chinese (5), Colchicine (4), Gases (10), Tocilizumab (15), Anakinra (4), IFN (11), Ig-based (4), Supplements (6)

### **REPOSITIONING CLINICAL TRIALS: ENROLLMENT**



• COVID-19 only

• Filtered out:

Observational studies; "not repurposed" (e.g., experimental); withdrawn; "Phase" N/A, or not applicable).

• Median: 275 patients



Avg(Enrollment) - +

#### https://clinicaltrials.gov/ct2/results?cond=COVID-19

Max

Median

#### **REPOSITIONING CLINICAL TRIALS: SUMMARY**







https://clinicaltrials.gov/ct2/results?cond=COVID-19

4/16/20 revision



https://clinicaltrials.gov/ct2/results?cond=COVID-19

4/16/20 revision

#### EXPERIMENTAL CLINICAL TRIALS: SUMMARY







https://clinicaltrials.gov/ct2/results?cond=COVID-19

4/23/20 revision

# TAKE HOME MESSAGE MOST COVID-19 INTERVENTIONAL **CLINICAL TRIALS ARE FOCUSED ON DRUG REPOSITIONING**

The majority study HCQ (76 out of 201)



4/23/20 revision

## HOW MANY DRUGS FOR REPOSITIONING?

	Drug Product				
	Forms		Type N Drug	Drugs	
	(Patents)*	Drug Products	Products	(Patents)*	RX Drugs **
<b>On-Patent</b>	12236 (4585)	1057	1057	<b>785</b> (738)	762
Off- Patent	22131	11874	1906	1454	1404
Discontinued	16963	11388	2801	<mark>680</mark> ^	n.a.
All Drugs	51330	22362	5042	2557	1828 (1258)

 Analysis based on the archived OrangeBook (2016 – 2019) and the latest Orange Book Data <u>Files</u> (OBDFs; EOBZIP\_2019\_10.zip content current as of: 10/18/2019) combined with the lists of Newly Added Patents and Delisted Patents

\* Number of Patents; \*\* Number of RX drugs as single ingredient; ^ 1833 total, only 680 discontinued

- Drug Product Forms = # of PIDs (all drug forms, routes, strengths etc)
- Drug Products = # of Application Numbers
- Type "N" Drug Products = # of Application Numbers of type "N". i.e., NDA (new drug application). The rest of the Drug products are "ANDA", i.e., abbreviated new drug applications.
- Drugs = # of (active) ingredients, i.e., actual drug (includes combinations)
- RX Drugs = # of drugs on prescription. The rest of the drugs are OTCs.

• Up to 1772 active ingredients may be eligible for "off patent" repurposing



S. Avram & T. Oprea, unpublished.

### DRUG REPOSITIONING INTELLIGENCE

#### Algorithmic Evaluation of Drug Repositioning Opportunities

Physico-Chemical Characteristics	Target & Ligand Based Virtual Screening	Disease-Target-Drug Associations	
Chemical fingerprints,	Shape & electrostatics	Disease & Drug target GO	
chemotypes, derivative	derived similarity or	Annotations from STRING,	
chemical descriptors	complementarity	GO, KEGG, other sources	
Solubility, Permeability, Distribution, relevant PK properties	3D models for intended targets, followed by molecular docking	Contigency Tables, e.g., Fisher test for associations augmented via Al/ML	
Metabolism, efflux transporters, toxicity	Multiple tautomers / protomers / conformers /	Prioritize novel disease- target associations for off-	
end-points	binding modes per protein	patent drugs	

Iteratively Compare, Integrate, Cross-check, Prioritize, Evaluate



Implement exact annotations for drug indications and off label uses. Mandate rigorous validation for computational models. Support community-based therapy-area specific research.

Oprea TI & Overington JP, Assay Drug Dev Technol. 2015 13:299-306 link

Turn DR into an international effort, preferably focused on diseases that lack cure



12/04/19 revision

#### DATA SOURCES AND METHODS

- Krogan et al., <u>preprint</u> (cleaned-up by Lars Jensen, PhD)
- <u>P-HIPSTER</u> predictions
- Metapath/XGBoost AIML predictions
- STRING analysis (<u>StringApp</u>, Cytoscape)
- <u>SmartGraph</u>



#### SARS-COV-2 INTERACTOME: TCLIN/TCHEM

- Tclin
- There are 169 significant (Fold change above 10) between 23 viral and 56 human proteins.
  Some occur multiple times, and are likely to be crucial in the way the virus subverts intracellular machinery.
- These are currently evaluated for potential repurposing.
- Tchem
- The Spike protein only had one significant (Fold change above 10) Tclin target, so we added 8 Tchem proteins for Spike and 6 for E-protein.
- Looking for drugs hitting these targets as well.
- *Note:* Giovanni Bocci is virtually screening viral targets for drug repositioning, and Cristian Bologa is virtually screening human targets for repositioning



# TCLIN AND ASSOCIATED DRUGS

PreyGene	Drug	Test
XPO1	selinexor	++++
IDH2	enasidenib	+++
GLA	migalastat	++
IMPDH2	mycophenolate mofetil	+++
UMPS	oteracil	++++
FDPS	zoledronic acid	++++
PSMB2	bortezomib	++++
NDUFA10	metformin	+++++
NDUFB10	metformin	
MT-ND1	metformin	
MT-ND3	metformin	
MT-ND5	metformin	
DNMT1	azacitidine	+++++
ITGB1	Natalizumab	++++
SLC29A1	Dipyridamole	+++
CRBN	lenalidomide	+++++

Moroxydine

From Wikipedia, the free encyclopedia

Moroxydine is an antiviral drug that was originally developed in the 1950s as an influenza treatment. It has potential applications against a number of RNA and DNA viruses.<sup>[1]</sup> Structurally moroxydine is a heterocyclic biguanidine.

It was reported in March 2014 that three kindergartens in two provinces of China had been found to be secretly dosing their students with moroxydine hydrochloride to try to prevent them from becoming ill. The kindergartens are paid only for the days that pupils attend and wanted to ensure that they maximised their earnings.<sup>[2]</sup>





Metformin is remarkably similar to moroxydine. From Krogan et al data, metformin targets ~20 proteins. Other potential mechanisms of action to be explored (based on the summary Table).



Tclin analysis by Tudor Oprea with input from John Overington

## PROTEIN KNOWLEDGE GRAPHS

Phenotype

Target

Disease

Endogenous ligand Drug



- IDG KMC2 seeks knowledge gaps across the five branches of the "knowledge tree":
- Genotype; Phenotype; Interactions & Pathways; Structure & Function; and Expression, respectively.
- We can use biological systems network modeling to infer novel relationships based on available evidence, and infer new "function" and "role in disease" data based on other layers of evidence
- Primary focus on Tdark & Tbio



O. Ursu, T Oprea et al., IDG2 KMC

## THE METAPATH APPROACH



Similar assertions or evidence form metapaths (white). Instances of metapath (paths) are used to determine the strength of the evidence linking a gene to disease/phenotype/function.

<u>G. Fu et al., *BMC Bioinformatics* 2016, 17:160</u> is an early example for drug discovery

- a meta-path is a path consisting of a sequence of relations defined between different object types (*i.e.*, structural paths at the meta level)
- Our metapaths encode typespecific network topology between the source node (e.g., Protein) and the destination node (e.g., Disease).
- This approach enables the transformation of assertions/evidence chains of heterogeneous biological data types into a ML ready format.



10/18/18 revision

## METAPATH/XGB00ST ML WORKFLOW



#### Transforming metapaths to ML features



Genes associated with ANY OMIM Phenotypic Series (PS) are the "train" subset divided into: ASSOCIATED with this PS  $\rightarrow$  RIGHT SIDE NOT ASSOCIATED with this PS  $\rightarrow$  LEFT SIDE Genes unassociated with ANY PS are the "test" subset  $\rightarrow$  LEFT SIDE.

#### Model dataset creation (details) Dataset here based on: OMIM phenotypic series







All datasets are merged, via R scripts, into a PostgreSQL. Python under development.

Graph embedding transforms evidence paths into vectors, converting data into matrices.

Input genes are positive labels. OMIM (not input) are negative labels (we prefer *true negatives* where possible).

XGBoost runs 100 models. The "median model" (AUC, F1) is then selected for analysis and prediction to avoid overfitting.

10/15/19 revision



#### METAPATH / XGBOOST MODEL INPUT

- P-HIPSTER: ACE2 (experimental) plus 25 other predicted proteins
- CD147 (experimental)
- 71 proteins (mass proteomic pull-down, Krogan et al paper)
- Total 98 positives
- Another 120 negatives from the Krogan paper
- 6 models built, based on variations of this input



ML Work by Praveen Kumar & Jeremy J Yang with input from Tudor Oprea

#### METAPATH / XGBOOST MODEL OUTPUT

- 986 proteins were predicted with "high confidence" by the 6 models
- 136 are predicted by 3 or more models.
- 99 of the 3x predicted proteins were Tbio/Tchem/Tclin
- These were used in combination with the input proteins plus the viral proteins from Krogan et al to examine the network models.

Tbio	
Tchem	
Tclin	

ML Work by Praveen Kumar & Jeremy J Yang





04/02/20 revision

#### SARS-COV-2 HUMAN PROTEIN INTERACTOME

SARS-CoV-2 proteins interact with multiple Tclin targets (blue).

*Less priority given to ATPand tubulin- type related targets.* 

Exploration in progress...



Analysis by Gergely Zahoranszky-Kohalmi



#### EARLY RESULTS

- HDAC2
- Interacts with SARS-CoV2 nsp5 and with SARS-CoV2-Spike.
- HDAC inhibitors: "HDIs have a long history of use in psychiatry and neurology as mood stabilizers and anti-epileptics. More recently they are being investigated as possible treatments for cancers, parasitic and inflammatory diseases."
- HDACs or HDAC inhibitors can be used to treat viral infections including coronavirus infections:
- <u>https://www.ncbi.nlm.nih.gov/pubmed/?term=28780424</u>
- <u>https://www.ncbi.nlm.nih.gov/pubmed/?term=23807710</u>



Preliminary Analysis by Lars Juhl Jensen

#### WAKE UP CALL ON HDAC2

- From: Willson, Tim [mailto:tim.willson[at]unc.edu]
  Sent: Thursday, March 26, 2020 4:38 PM
  Reference 1 says in its abstract "Surprisingly, the antiviral activity of U18666A was suppressed by the histone deacetylase inhibitor (HDACi), Vorinostat"
- Reference 2 says in the abstract "Not surprisingly, viruses have evolved a wide array of mechanisms to subvert HDAC functions."
- I have not read the papers, but sounds like an HDAC inhibitor is likely to <u>promote</u> replication of the virus
- Viruses appear to HDAC activity so they can initiate their own replication. So the association is real, it just favors viruses. Tim suggested the reverse effect by blocking HATs, histone acetyl transferases.



#### INDIRECT(?) CONFIRMATION

- Pracinostat, vorinostat and panobinostat appear to *accelerate* the virus-induced killing process
- Alternatively, these drugs directly kill VEROE6 cells.
- This preliminary finding encourages us to further pursue the histone deacetylase hypothesis



HATS

EP300 O09472 Tchem HAT1 O14929 Tbio KAT2A **O**92830 Tbio KAT2B **O**92831 Tchem KAT5 **O**92993 Tchem KAT6A O92794 Tbio KAT6B O8WYB5 Tbio KAT7 **O**95251 Tbio KAT8 **O9H7Z6** Tchem **NAA60 O9H7X0** Tbio RBBP7 O16576 Tbio **CREBBP** O92793 Tchem ATF2 P15336 Tbio TAF1 P21675 Tchem **NAA40 O86UY6** Tbio NCOA1 **O**15788 Tchem NCOA3 **O9Y6O9** Tbio



AXIN1

- One of the HATs, RBBP7, is consistently overexpressed when interacting w/ SARS-CoV2. Two others, HAT1 and NAA40, are relatively under-expressed (data from Krogan et al)
- No other HATs are on the list.
- Graph (right): "chemicalizing" the HATs network using <u>SmartGraph</u>

**PIDG** 

SmartGraph Analysis by Gergely Zahoranszky-Kohalmi

04/02/20 revision

#### DRUG-TARGET INTERACTIONE MODULATING HATS



## VIRTUAL SCREENING AGAINST SARS-COV-2 VIRAL

No. Name	Pocket	Туре	Priority	Drug	CMax (uM)	PDB Template
l chloroquine		active ligand	1		0.01	
2 remdesivir (pro-drug)		active ligand	1	tegaserod	0.01	5E6J
32-O-MTase	4A	binding site	1	triamterene	0.33	6NUR
4 3CLpro	1B	binding site	1		<b>D</b> 41	
5 3CLpro	2AB	binding site	1	meloxicam	5.41	5E6J   6NUK
6 3CLpro	3AB	binding site	1	ibuprofen	295.71	5E6J 6NUR
7 Helicase	9B	binding site	1	- naprovon	100 01	FEGLENITE
8 Helicase	10B	binding site	1	napioxen	400.24	SEOJONOK
9 Helicase	18B	binding site	1	hydrochlorothiazide	0.25	6NUR
10N7-MTase	13D	binding site	1	baclofon	0.75	FEAT
11 RdRp	6A	binding site	1	Dacioleli	0.15	5E0J
12 RdRp	17A	binding site	1	trimethoprim	4.13	5E6J 6NUR
13 remdesivir triphosphate	5A	active ligand	2	- ethambutol	17 13	SFELLENIIR
142-O-MTase	5A	binding site	2	emanduloi	11.10	
15 Helicase	11B,12B	binding site	2	cidofovir	70.20	5E6J 6NUR
16 N7-MTase	14D	binding site	2	ivazomih	0 17	5761
17 NendoU	15A	binding site	2	IXazoIIIID	0.11	510
18 RdRp	7A	binding site	2	safinamide	3.31	5E6J 6NUR
19 RdRp	8A	binding site	2	avibactam	55.04	6NUR

Start: 3,981 small molecules drugs from <u>DrugCentral</u>. Enumerated tautomeric and protomeric forms available at pH 7.4 with minimum abundance of 25% (total, 6057 structures). These were *initially* docked into the main SARS-CoV targets (selected drugs on the right):

- 5E6J: SARS Coronavirus Papain-like Protease
- 6NUR: SARS-CoV nsp12 polymerase Work in progress

Work by Giovanni Bocci

TARGETS



#### FURTHER EXPERIMENTAL RESULTS

April 17, 2020

Chloroquin

NUUUUUUUUU

	Prediction	VeroE6		
Drug Name	Source	-CoV2	BDDCS	<del>โอ</del> <sup>1.5</sup> ๆ เ
Lenalidomide	MP/XGb	x	4	q L
Mycophenolate	MP/XGb	x	2	
Fasudil HCl	Network	xx	1	
Safinamide Mesylate	SBVS-CoV2	xx	2	יחושההוה הוה במחדרות בישו
Meloxicam	SBVS-CoV2	x	2	
Ibuprofen	SBVS-CoV2	x	2	ve 0.5−1 11111111111111111111111111111111111
Ethambutol HCl	SBVS-CoV2	x	3	$\tilde{\boldsymbol{\Sigma}}$
Trimethoprim	SBVS-CoV2	xx	3	
Triamterene	SBVS-CoV2	x	2	<u>о <sub>0.0</sub> , , , , , , , , , , , , , , , , , , ,</u>

• Currently pursuing further experiments on these drugs, plus experiments on *combinations.* 

DMSO uninfecte DMSO infecte

#### EXPERIMENTAL CONFIRMATION

- "If you have one clock, you know what time it is. Having two clocks, you can no longer tell time precisely."
- Many experimental papers have been published. The overlap between them is not exactly encouraging.
- Comparing multiple experiments, we start to understand the need for confirmation, the essence of science.
- The combination of computational and experimental screening is more likely to lead to practical therapeutic solutions.



# TAKE HOME MESSAGE DRUG REPOSITIONING FOR COVID-19 IS AN ITERATIVE EXERCISE

Experiments need to inform future experiments Multiple approaches lead to different drugs



4/23/20 revision