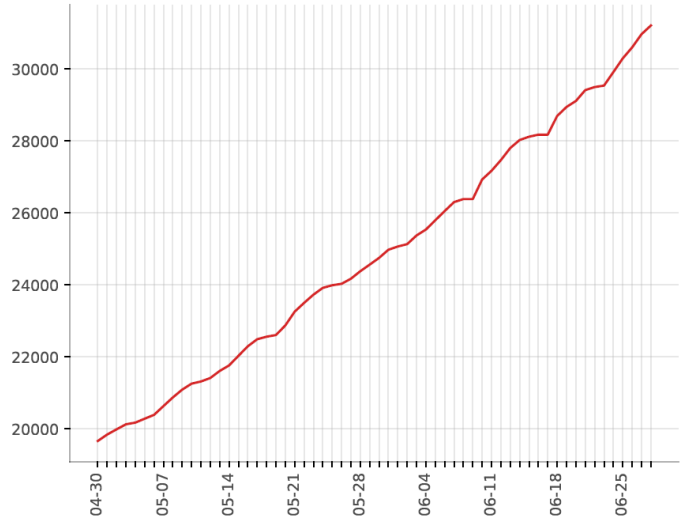
COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation

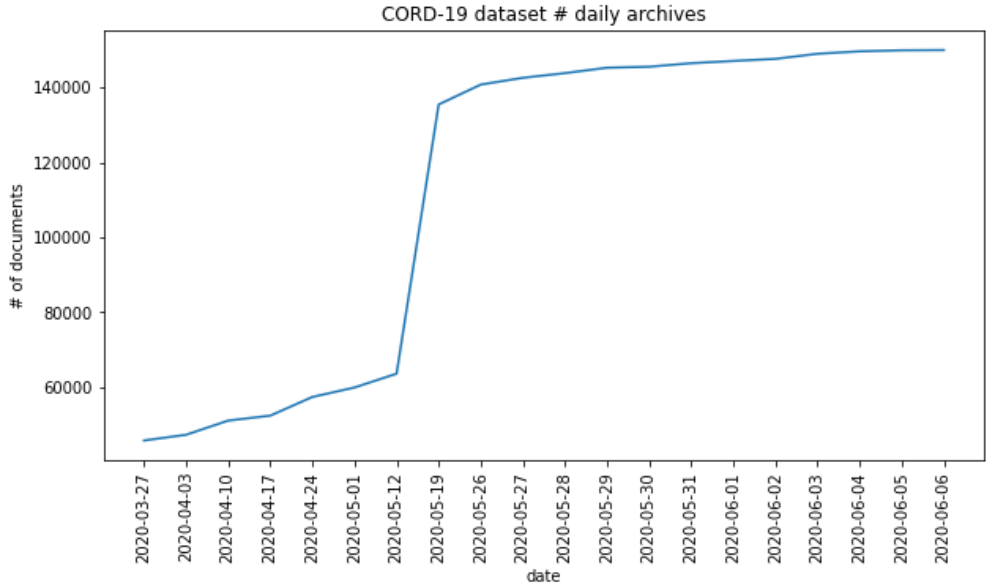
Qingyun Wang[1], Manling Li[1], Xuan Wang[1], Nikolaus Parulian[1], Guangxing Han[2], Jiawei Ma[2], Jingxuan Tu[3], Ying Lin[1], Haoran Zhang[1], Weili Liu[1], Aabhas Chauhan[1], Yingjun Guan[1], Bangzheng Li[1], Ruisong Li[1], Xiangchen Song[1], Yi R. Fung[1], Heng Ji[1], Jiawei Han[1], Shih-Fu Chang[2], James Pustejovsky[3], Jasmine Rah[4], David Liem[5], Ahmed Elsayed[6], Martha Palmer[6], Clare Voss[7], Cynthia Schneider[8], Boyan Onyshkevych[9]

[1]University of Illinois at Urbana-Champaign [2]Columbia University [3]Brandeis University
[4]University of Washington [5]University of California, Los Angeles [6]Colorado University
[7]Army Research Lab [8]QS2 [9]Defense Advanced Research Projects Agency

Challenges on Digesting COVID-19 Scientific Literature

- Practical progresses at combating COVID-19 highly depend on effective transmission, assessment and extension of research results
- Quantity:
  - **2.7K** new papers per day
  - As of June 13, 2020, there are at least **140K** papers about coronavirus
- Quality:
  - Given the rapid publications of preprints without peer reviews, many research results are **redundant**, **complementary** or even **conflicting** with each other





Temperature, humidity, and latitude analysis to predict potential spread and seasonality for COVID-19 **Positive**

Mohammad M. Sajadi, MD,[1,2] Parham Habibzadeh, MD,[3] Augustin Vintzileos, PhD,[4] Shervin Shokouhi, MD,[5] Fernando Miralles-Wilhelm, PhD,[6-7] Anthony Amoroso, MD[1,2]

No Association of COVID-19 transmission with temperature or UV radiation in Chinese cities **Negative**

[†]Dr. Yao, Ms. Pan, Ms. Liu and Dr. Meng contributed equally to this letter.

Challenges on Digesting COVID-19 Scientific Literature

- Knowledge Bottlenecks in Clinical Trials for Drug re-purposing
  - Mainly rely on symptoms: consider drugs that can treat diseases with similar symptoms
  - Too many drug candidates
  - Too much misinformation from multiple countries
  - Too costly to test all drugs, and difficult to decide threshold to measure success (Is it good enough to if 65% patients have reduced symptoms?)
- What doctors need: a reliable ranked list of drugs with detailed knowledge-level (chemicals/genes) evidence

Drug Repurposing Report for hydroxychloroquine

Section 1. Current indication: what is the drug class? What is it currently approved to treat?

Drug class = antimalarials

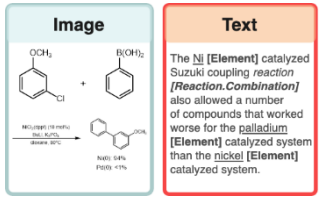
PMID	Sentence
32278373 PMC7146712	With no "adequate, approved and available" alternative, the US Food and Drug Administration (FDA) is allowing the use of the antimalarial drugs hydroxychloroquine and chloroquine to treat coronavirus disease 2019 (COVID-19).
32385749 PMC7207983	The World Health Organization (WHO) launched the SOLIDARITY trial on 20 March which includes already approved drugs for other diseases repurposed to treat COVID-19. Mechanisms of action of these drugs are already known as well as their toxicity. We highlight the potential risks of combining drugs with established macular toxicity that are hydroxychloroquine and ritonavir, as actually performed in our Lombardia region, the heart of COVID-19 Italian break-out.
32267979 PMC7262068	Pharmacotherapy for COVID-19 is limited and treatment remains primarily supportive. At the time of writing, no COVID-19-specific pharmacotherapies were approved by the FDA (on March 29, the FDA authorized emergency use of chloroquine and hydroxychloroquine). Nevertheless, the gravity of the situation has led to a flurry of anecdotal pharmacotherapy approaches in the ICU.
32511331 PMC7239066	There is interest in the use of chloroquine/hydroxychloroquine (CQ/HCQ) and azithromycin (AZT) in COVID-19 therapy. Employing cystic fibrosis respiratory epithelial cells, here we show that drugs AZT and ciprofloxacin (CPX) act as acidotropic lipophilic weak bases and confer in vitro effects on intracellular organelles similar to the effects of CQ. These seemingly disparate FDA-approved antimicrobials display a common property of modulating pH of endosomes and trans-Golgi network.
32299202 PMC7195984	Chloroquine (CQ) and hydroxychloroquine (HCQ) are 4-aminoquinoline derivatives that are approved by the U.S. Food and Drug Administration (FDA) for the treatment of malaria, systemic lupus erythematosus, rheumatoid arthritis (RA) and decades of experience in use of these disorders. They are also used in Q fever and porphyria cutanea tarda. HCQ has a better side effect profile than CQ and is strongly recommended for the long-term treatment of lupus unless the occurrence of a severe side effect.
32304645 PMC7158822	Additional active clinical trials involve the use of drugs approved for different therapeutic indications. This is the case, for example, for: (i) the FDA-approved antimalarial drugs chloroquine and hydroxychloroquine, owing to their ability to interfere with basic cellular pathogenetic mechanisms; and (ii) monoclonal antibodies against interleukin-6 receptor (anti-IL-6R) which might be helpful in reducing abnormal inflammatory response upon cytokine storm, thus improving organ functions in COVID-19 patients. This 'recycling' strategy based on the re-use of approved drugs is commonly referred to as drug repurposing and is largely successful, as demonstrated by examples of repurposing treatments in cancer and other human diseases. Drug repurposing is a modern therapeutic strategy that substantially reduces the risks of drug development and costs. In this emergency, it shortens the time gap between the identification of a potentially useful drug and the treatment of the patient owing to the availability of large amounts of safety, tolerability, pharmacokinetic, pharmacodynamic and clinical data on the existing drug.
32306836 PMC7196923	In the view of urgency and current need, to reduce the cost, time and risks of the drug development process, scientists are involved in reusing already approved drug candidates to test in COVID-19 patients. In a short time, the response of the scientific community is such that involve enormous efforts to develop a novel therapy and treatment. For example, Chloroquine and Hydroxychloroquine, old drugs, have been used to treat malarial, rheumatoid arthritis, lupus and sun allergies for more than sixty years. The activity of hydroxychloroquine on viruses is probably same as that of chloroquine since the mechanism of the action of these two molecules is identical. Chloroquine as an antimalarial and autoimmune disease drugs has shown a synergistically enhancing effect as antiviral drugs in vivo studies (Savarino et al.,; Yan et al.,).

Our Goals

DATA

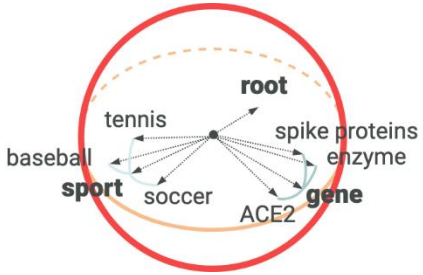
SEMANTICS

KNOWLEDGE BASE



Scientific Literature

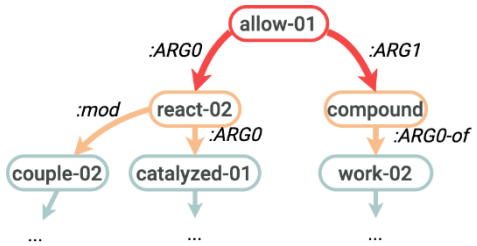
Chemical Ontology & Existing Databases



Hierarchical Spherical Embedding

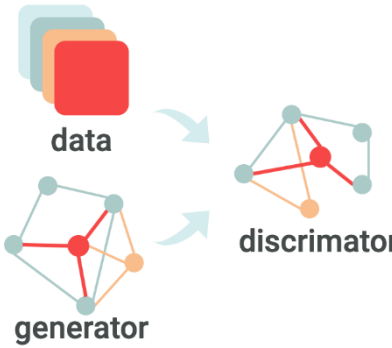
KB concept of Ni	
Name: Nikel	Element category: transition metal
Chemical Symbol: Ni	Appearance: lustrous, metallic, and silver with a gold tinge
Atomic Number: 28	
Type: chemical element	

Ontology Enriched Text Embedding



Cross-media Structured Semantic Representation

Graph neural networks
Joint entity/relation/event extraction
and ontology construction



Generative Adversarial Networks for Data Augmentation and Distant Supervision

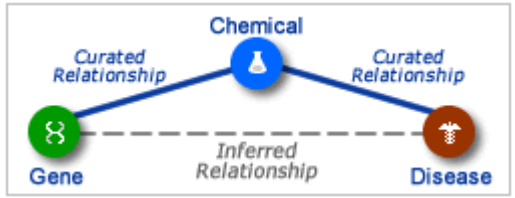
Multimedia Knowledge Base

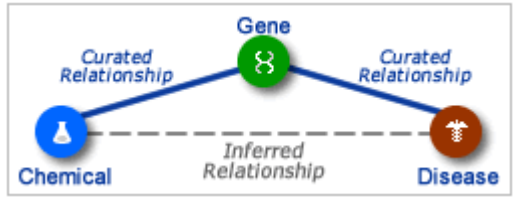
Multimedia Search and Summarization

[Daily paper feeds, KGs and drug repurposing reports at http://blender.cs.illinois.edu/covid19](http://blender.cs.illinois.edu/covid19)

Coarse-grained Text Knowledge Extraction

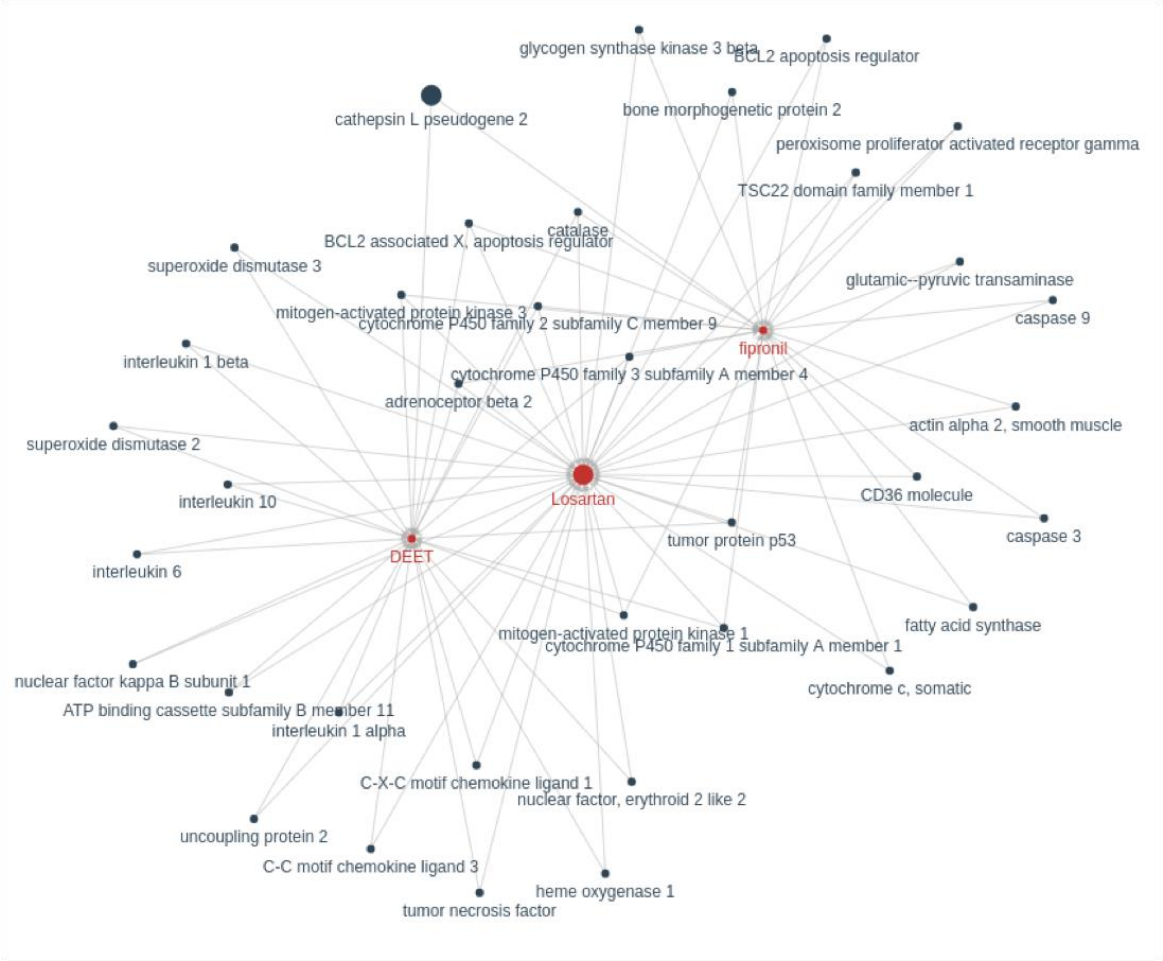
- Entity Extraction + Entity Linking:
  - Extract entities from unstructured texts, link entity mentions to external biomedical ontologies including Comparative Toxicogenomics Database (CTD) (Davis et al., 2016) and obtain a Medical Subject Headings (MeSH) Unique ID for each mention.
- Relation Extraction:
  - Extract 133 relation types including Gene–Chemical–Interaction Relationships, Chemical–Disease Associations, Gene–Disease Associations, Chemical–GO Enrichment Associations and Chemical–Pathway Enrichment Associations
- Event extraction:
  - Extract 13 Event types and the roles of entities involved in these events, including Gene expression, Transcription, Localization, Protein catabolism , Binding, Protein modification, Phosphorylation , Ubiquitination, Acetylation, Deacetylation, Regulation, Positive regulation, Negative regulation





Coarse-grained Text Knowledge Extraction

- Knowledge extraction from 25,534 COVID-19 papers to construct knowledge graphs
- Current KG:
  - 50,864 Gene nodes, 7,230 Disease nodes, 9,123 Chemical nodes, 1,725,518 chemical-gene links, 5,556,670 chemical-disease links, and 77,844,574 gene-disease links
- The figure shows an example of the constructed KG from multiple papers connecting a candidate drug in COVID-19 (Losartan) and a gene related to coronavirus (cathepsin L pseudogene)
  - The red nodes represent chemicals, grey nodes represents genes, and edges represent gene-chemical relations
- Experiments on 186 documents with 12,916 sentences manually annotated by domain experts show that our method achieves **83.6**% F-score on node extraction and **78.1**% F-score on link extraction.



Fine-grained Text Knowledge Extraction

- Fine-grained entity extraction (CORD-NER) for 75 entity types (Wang and Han, 2020) such as coronaviruses, viral proteins, evolution, materials, immune responses, etc
- CORD-NER relies on distantly- and weakly-supervised methods with no need for expensive human annotation.
- Its entity annotation quality surpasses SciSpacy (up to 93.95% F-score, over 10% higher on the F1 score based on a sample set of documents), a fully supervised BioNER tool.
- So we are able to answer questions including fine-grained entities such as “Which amino acids in glycoprotein are most related to the CHEMICAL?”

Spacy (General NER):

Angiotensin-converting enzyme 2 **CARDINAL** (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target.

SciSpacy (Biomedical NER):

Angiotensin-converting enzyme 2 **GENE_OR_GENOME** (ACE2 **GENE_OR_GENOME**) as a SARS-CoV-2 receptor **GENE_OR_GENOME** : molecular mechanisms and potential therapeutic target.

Ours:

Angiotensin-converting enzyme 2 **GENE_OR_GENOME** (ACE2 **GENE_OR_GENOME**) as a SARS-CoV-2 **CORONAVIRUS** receptor: molecular mechanisms and potential therapeutic target.

Spacy (General NER):

A phylogenetic analysis [3 **CARDINAL** , 4 **CARDINAL**] found a bat origin for the SARS-CoV-2.

SciSpacy (Biomedical NER):

A phylogenetic analysis [3, 4] found a bat origin for the SARS-CoV-2 **SIMPLE_CHEMICAL** .

Ours:

A phylogenetic **EVOLUTION** analysis [3 **CARDINAL** , 4 **CARDINAL**] found a bat **WILDLIFE** origin for the sars_cov_2 **CORONAVIRUS** .

Fine-grained Text Knowledge Extraction

- Here is a figure for results on part of a COVID-19 paper

Angiotensin-converting enzyme 2 ***GENE_OR_GENOME*** *(* *ACE2* ***GENE_OR_GENOME*** *) as a SARS-CoV-2* ***CORONAVIRUS*** *receptor* ***CHEMICAL****: molecular mechanisms and potential therapeutic target.*

SARS-CoV-2 **CORONAVIRUS** has been sequenced [3] . A phylogenetic **EVOLUTION** analysis [3 , 4] found a bat **WILDLIFE** origin for the SARS-CoV-2 **CORONAVIRUS** . There is a diversity of possible intermediate hosts **NORP** for SARS-CoV-2 **CORONAVIRUS** , including pangolins **WILDLIFE** , but not mice **EUKARYOTE** and rats **EUKARYOTE** [5] . There are many similarities of SARS-CoV-2 **CORONAVIRUS** with the original SARS-CoV **CORONAVIRUS** . Using computer modeling , Xu et al **PERSON**. [6] found that the spike proteins **GENE_OR_GENOME** of SARS-CoV-2 **CORONAVIRUS** and SARS-CoV **CORONAVIRUS** have almost identical 3-D structures in the receptor binding domain that maintains Van der Waals forces **PHYSICAL_SCIENCE** . SARS-CoV spike proteins **GENE_OR_GENOME** has a strong binding affinity **DISEASE_OR_SYNDROME** to human ACE2 **GENE_OR_GENOME** , based on biochemical interaction studies and crystal structure analysis [7] . SARS-CoV-2 **CORONAVIRUS** and SARS-CoV spike proteins **GENE_OR_GENOME** share identity in amino acid sequences and , importantly, the SARS-CoV-2 **CORONAVIRUS** and SARS-CoV spike proteins **GENE_OR_GENOME** have a high degree of homology [6, 7] . Wan et al **PERSON**. [4] reported that residue 394 **CARDINAL** (glutamine **CHEMICAL**) in the SARS-CoV-2 **CORONAVIRUS** receptor-binding domain ….

Image Processing and Cross-media Entity Grounding

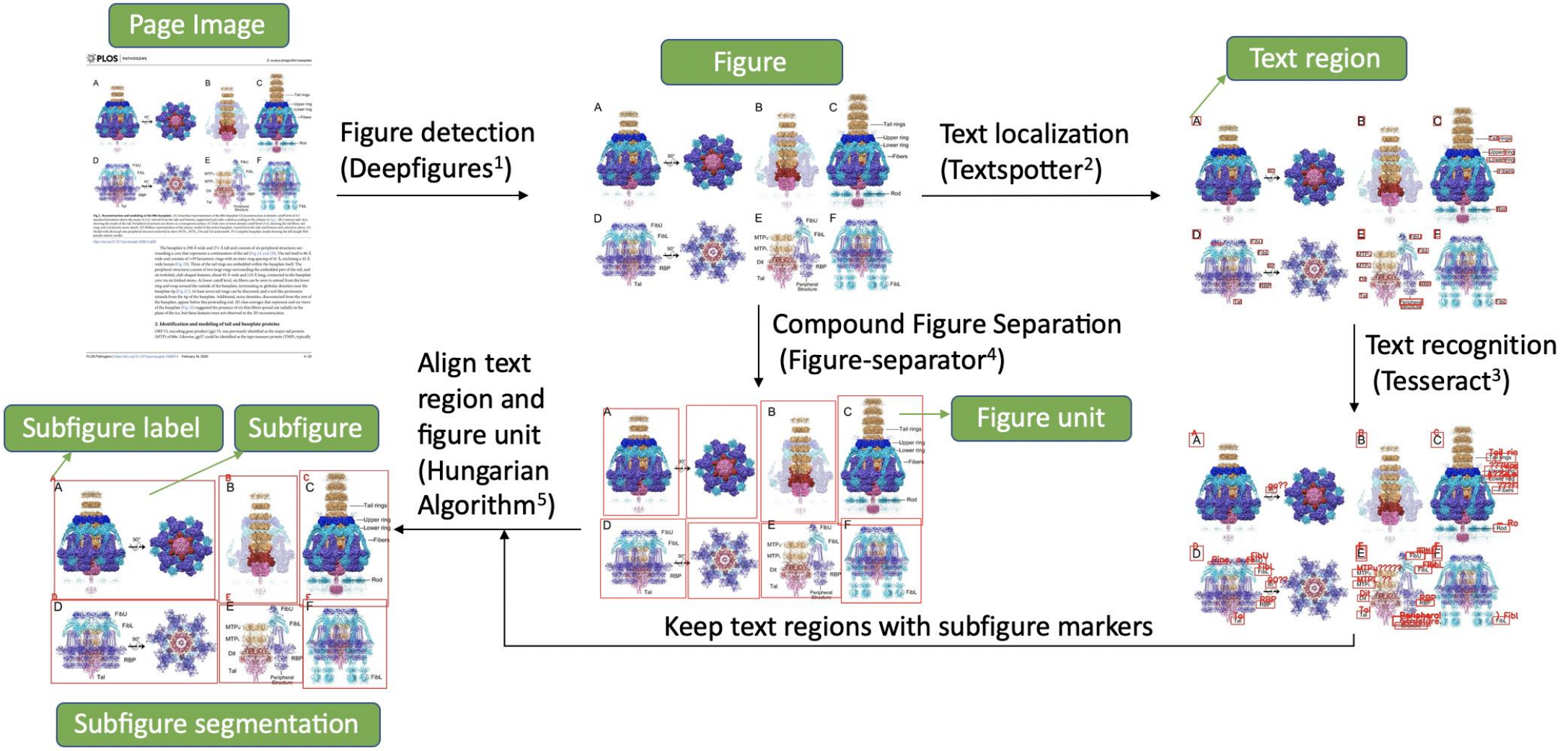
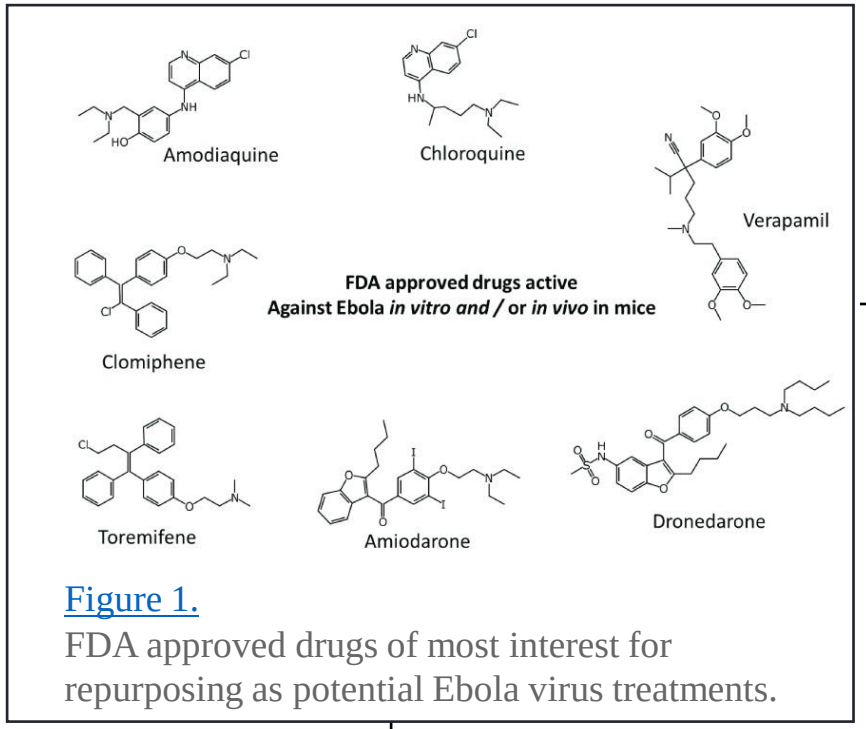
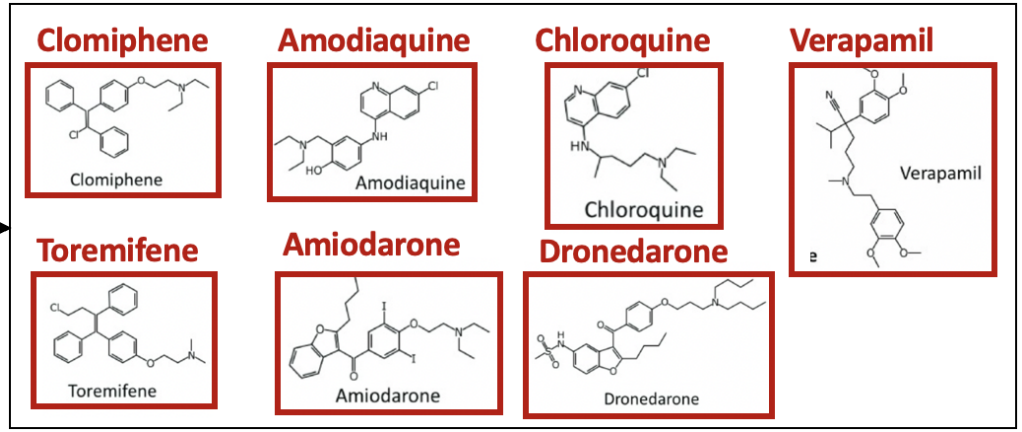


Image Processing and Cross-media Entity Grounding

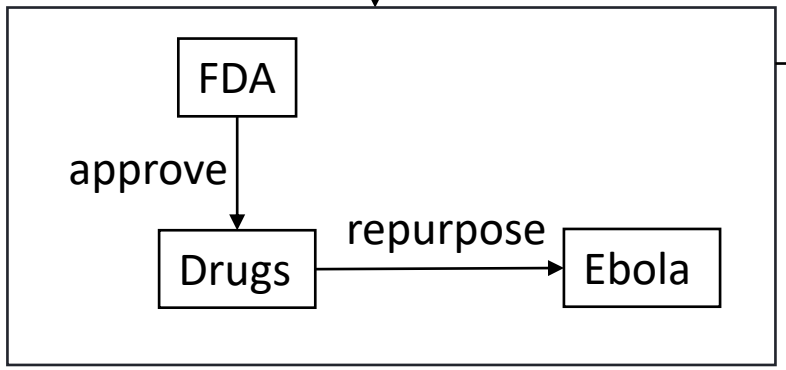


[Figure 1.](#)
FDA approved drugs of most interest for repurposing as potential Ebola virus treatments.

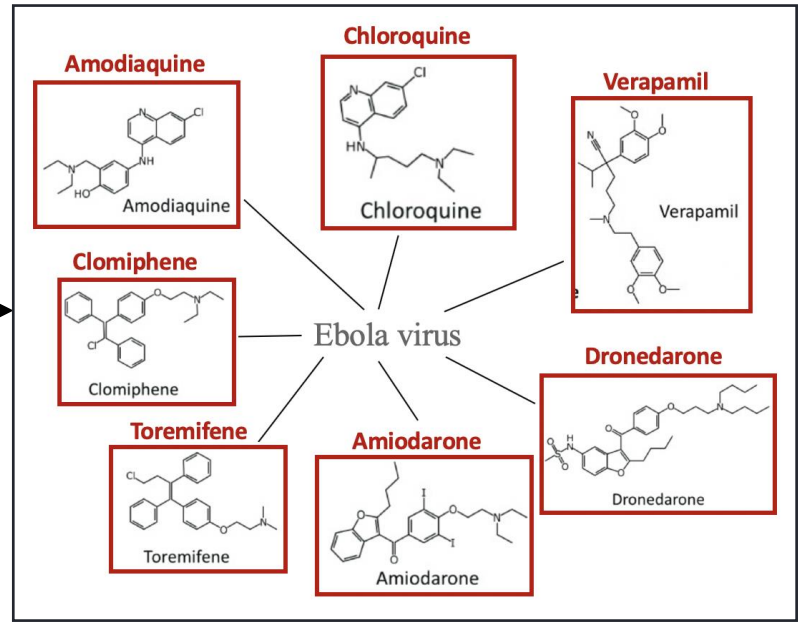
Entity Grounding for Drug Molecular Structure Image



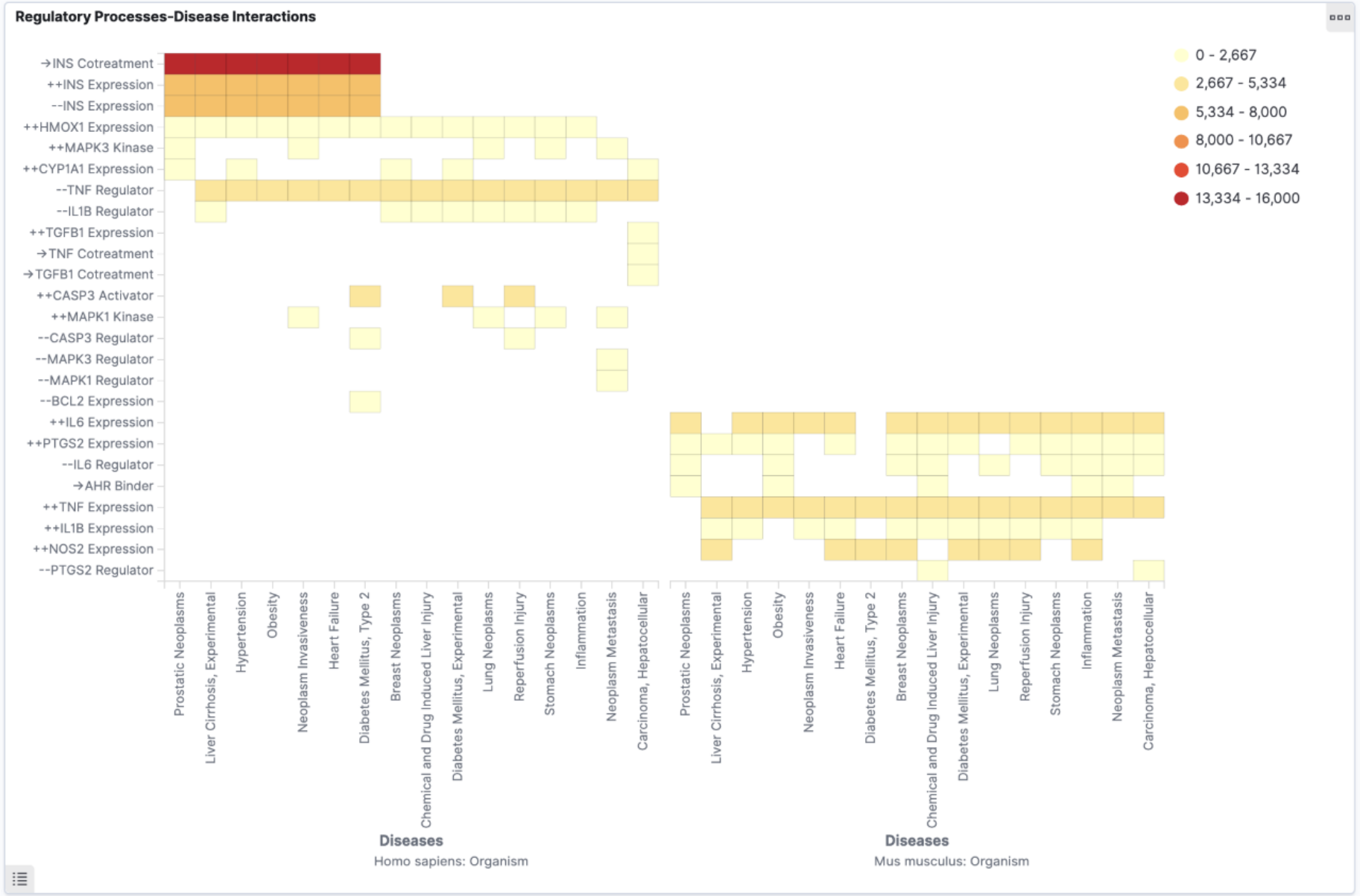
KG from caption text



Multimedia Knowledge Graph Expansion



Knowledge Graph Semantic Visualization

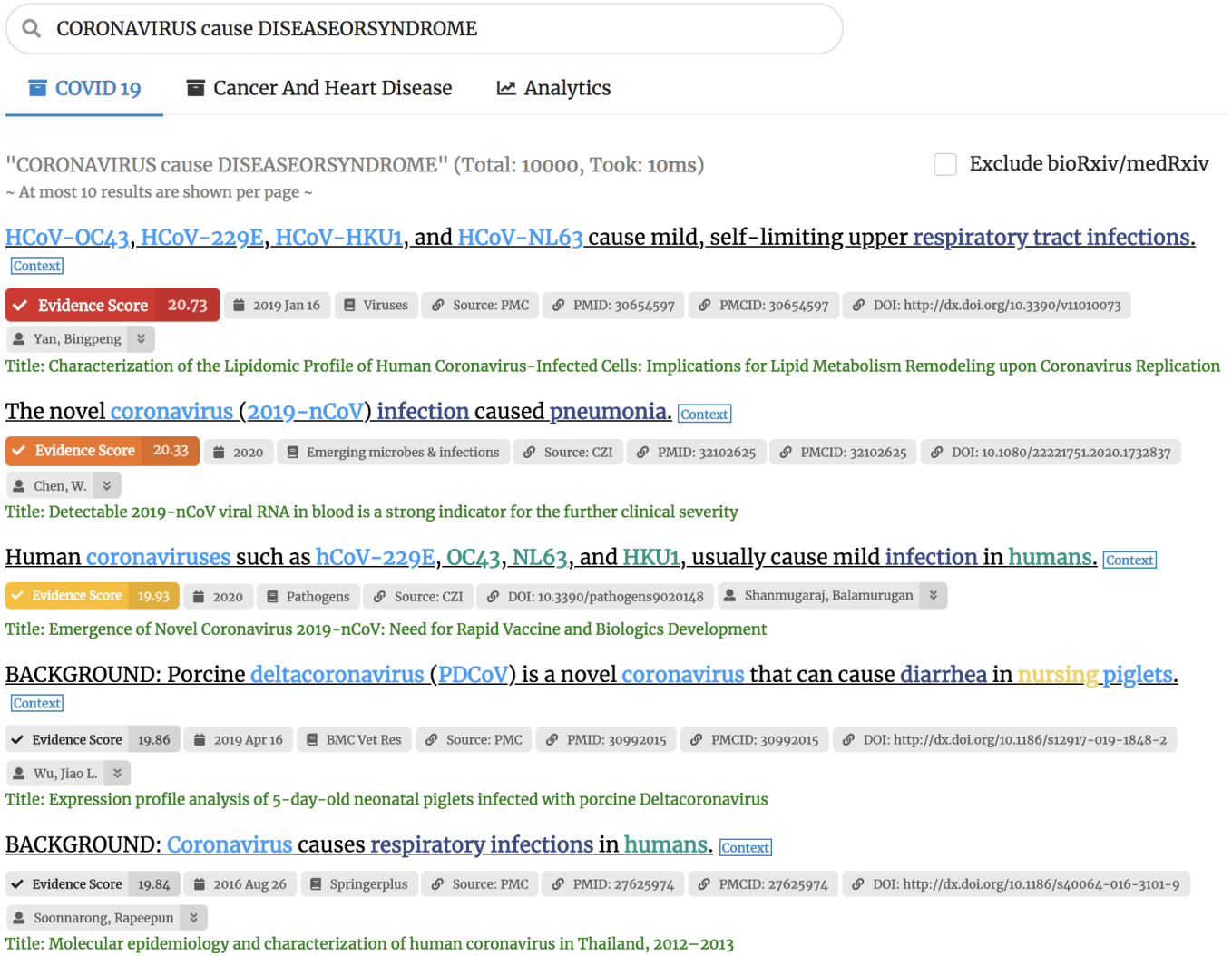


Knowledge-driven Question Answering

- Limitations of state-of-the-art question answering
  - Fully rely on Word-level or sentence-level semantic meaning matching
  - Questions are limited to non-experts (e.g., “Corona Virus Update?”) or too high-level (e.g., “What is known about transmission, incubation, and environmental stability?”)
- What we need: install a scientific brain (knowledge base) for QA
- Preliminary Results

Question	# of Answers	Example Answers
Which genes are related to COVID-19?	687	AP2 associated kinase 1, myeloperoxidase, thioredoxin
Which chemicals are related to COVID-19?	3,142	acetoacetic acid, Chlorine, Zymosan
Which diseases are the most similar to COVID-19?	4	Enteritis, Transmissible, of Turkeys; Feline Infectious Peritonitis; Gastroenteritis, Transmissible, of Swine; Severe Acute Respiratory Syndrome
Which genes are related to COVID-19 that can be transferred from its similar diseases?	2,168	DEK proto-oncogene, neclear receptor corepressor 1
Which chemicals are related to COVID-19 that can be transferred from its similar diseases?	327	Ampicillin, Quercetin, Zoledronic Acid

EvidenceMiner with Query: “CORONAVIRUS cause DISEASEORSYNDROME”



Online demo: https://evidenceminer.com/

A Case Study on Drug Repurposing Report Generation

- Target Chemicals/Genes from DARPA Biologists
- BM1_00870 BM1_06175 BM1_16375 BM1_17125 BM1_22385 BM1_30360 BM1_33735 BM1_56245 BM1_56735 CATB-10270 CATB-1418 CATB-1674 CATB-16A CATB-16D2 CATB-1852 CATB-1874 CATB-2744 CATB-3098 CATB-348 CATB-3483 CATB-5880 CATB-84 CATB-912 CATD CATHY CATK CATL CATL-LIKE CTS12 CTS3 CTS6 CTS7 CTS7-PS CTS8 CTS8L1 CTS8-PS CTSA CTSA.L CTSB CTSBA CTSBB CTSB.L CTSB-PS CTSB.S CTSC CTSC.L CTSC.S CTSD CTSD2 CTSD.S CTSE CTSEAL CTSE.L CTSE.S CTSF CTSF.L CTSG CTSH CTSH.L CTSH-PS CTSJ CTSK CTSK1 CTSK.L CTSL CTSL.1 CTSL3 CTSL3P CTSLA CTSLB CTSLL CTSL.L CTSLL3 CTSLP1 CTSLP2 CTSLP3 CTSLP4 CTSLP6 CTSLP8 CTSM CTSM-PS CTSM-PS2 CTSO CTSO.L CTSQ CTSQL2 CTSR CTSS CTSS1 CTSS.2 CTSS2.1 CTSS2.2 CTSSL CTSS.L CTSS.S CTSV CTSV.L CTSW CTSW.L CTSZ CTSZ.L CTSZ.S LOAG_18685 SMP_013040.1 SMP_034410.1 SMP_067050 SMP_067060 SMP_085010 SMP_085180 SMP_103610 SMP_105370 SMP_158410 SMP_158420 SMP_179950 TSP_01409 TSP_02382 TSP_02383 TSP_03306 TSP_07747 TSP_10129 TSP_10493 TSP_11596 LMAN1 LMAN1L LMAN1.L LMAN1.S LMAN2 LMAN2L MBL1P MBL2 ACE2 FURIN TMPRSS2

- Since human reports follow a template of 10 sections, let’s try to automate the report generation!
- Pick three drugs for case study: Benazepril, Losartan, Amodiaquine
- Report (64 pages) at: http://blender.cs.illinois.edu/covid19/DrugRe-purposingReport_V2.0.docx

Section 1: Current indication: what is the drug class? What is it currently approved to treat?

- Example output for Benazepril

Drug Class = angiotensin-converting enzyme (ACE) inhibitors

It is currently approved to treat:

Disease	**Hypertension**
PMID, PMCID	Evidence Sentences
32314699 PMC7253125	Past medical history was significant for hypertension, treated with amlodipine and benazepril, and chronic back pain.
32081428 PMC7092824	On the other hand, many ACE inhibitors are currently used to treat hypertension and other cardiovascular diseases. Among them are captopril, perindopril, ramipril, lisinopril, benazepril, and moexipril.

Disease	**Dogs with Congestive Heart Failure (CHF)**
PMID	Evidence Sentences
31254308 PMC6639469	9-year-old male neutered Cocker Spaniel with severe CHF receiving furosemide, benazepril, hydrocodone, sildenafil, and pimobendan;

(2) path connecting drug and disease in KG

Disease	**Covid-19**
PMID, PMCID	**Evidence Sentences**
32081428, PMC7092824	By using a molecular docking approach, an earlier study identified N-(2-aminoethyl)-1 aziridine-ethanamine as a novel ACE2 inhibitor that effectively blocks the SARS-CoV RBD-mediated cell fusion. This has provided a potential candidate and lead compound for further therapeutic drug development. Meanwhile, biochemical and cell-based assays can be established to screen chemical compound libraries to identify novel inhibitors. On the other hand, many ACE inhibitors are currently used to treat hypertension and other cardiovascular diseases. Among them are captopril, perindopril, ramipril, lisinopril, benazepril, and moexipril. Although these drugs primarily target ACE, a homolog of ACE2 with 42% sequence identity and 61% sequence similarity in the catalytic domain, they may be effective toward ACE2 as well.

Section 2: Molecular structure (symbols desired, but a pointer to a reference is also useful)

Benazepril ($C_{24}H_{28}N_2O_5$): an angiotensin converting enzyme (ACE) inhibitor, used alone or together with other medicines to treat high blood pressure (hypertension).

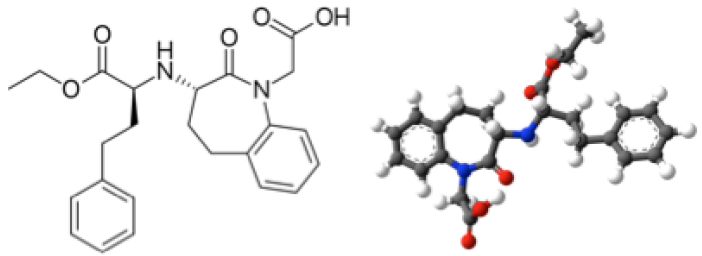


Figure: Molecular structure of Benazepril (src: wikipedia)

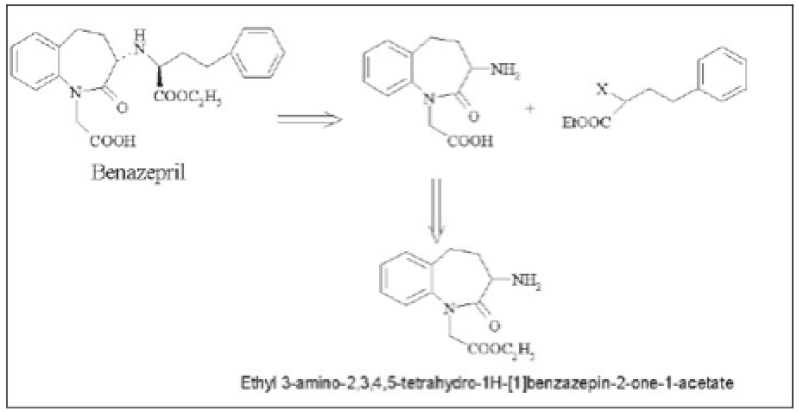


Figure: Retrosynthesis of benazepril (src: PMC3003176 Fig.1)

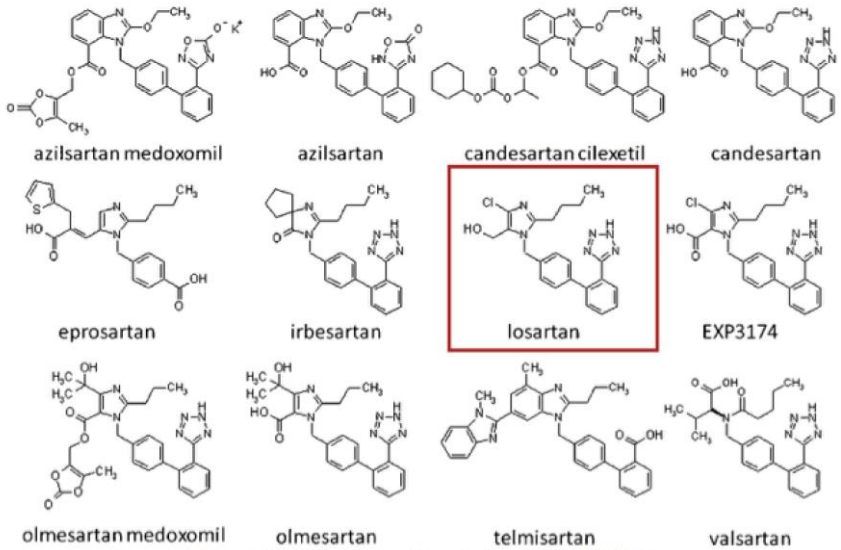


Figure: Molecular structure of Losartan (src: PMID 17043154)

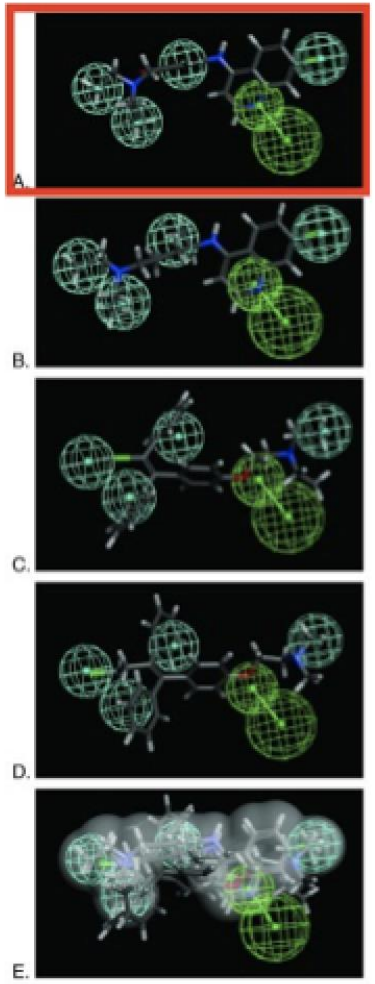


Figure: Src: PMC4304229 Fig 1. Pharmacophore based on 4 hits.

Section 3: Mechanism of action i.e. inhibits viral entry, replication, etc (w/ a pointer to data)

PMID	Chemical (ID)	Gene (ID)	Interaction	InteractionActions
18713951	benazepril (MESH:C044946)	SLC15A1 (6564)	benazepril inhibits the reaction [SLC15A1 protein results in increased uptake of glycylsarcosine]	decreases^reaction, increases^uptake
19018797	benazepril (MESH:C044946)	RELA (5970)	benazepril results in decreased phosphorylation of RELA protein	decreases^phosphorylation
20821936 16732983	benazepril (MESH:C044946)	MMP2 (4313)	benazepril inhibits the reaction [Streptozocin results in decreased expression of MMP2 mRNA]	decreases^expression, decreases^reaction
1179862711 8679511619 1423171771 3819018797	benazepril (MESH:C044946)	TGFB1 (7040)	benazepril results in decreased expression of TGFB1 mRNA	decreases^expression
1150106211 7986271901 8797	benazepril (MESH:C044946)	AGT (183)	benazepril results in decreased expression of AGT protein modified form	decreases^expression
1113670087 63405	benazepril (MESH:C044946)	NPPA (4878)	benazepril results in decreased expression of NPPA mRNA	decreases^expression
1531534116 364833	benazepril (MESH:C044946)	ACE (1636)	benazepril results in decreased activity of ACE protein	decreases^activity
1265232712 8489191555 5355157937 8715788353 1100783115 498266	benazepril (MESH:C044946)	ACE (1636)	ACE gene polymorphism affects the susceptibility to benazepril	affects^response to substance
16635409	benazepril (MESH:C044946)	SMAD2 (4087)	benazepril affects the expression of SMAD2 protein	affects^expression
1673298320 821936	benazepril (MESH:C044946)	TIMP2 (7077)	benazepril inhibits the reaction [Streptozocin results in increased expression of TIMP2 mRNA]	decreases^reaction, increases^expression
20671225	benazepril (MESH:C044946)	APOB (338)	Fluvastatin promotes the reaction [[Valsartan co-treated with benazepril] results in decreased expression of APOB protein]	affects^cotreatment, decreases^expression, increases^reaction
16775501	benazepril (MESH:C044946)	TNF (7124)	[benazepril co-treated with Amlodipine] results in decreased expression of TNF protein	affects^cotreatment, decreases^expression
16635409	benazepril (MESH:C044946)	SMAD2 (4087)	[benazepril co-treated with Irbesartan] affects the expression of SMAD2 mRNA	affects^cotreatment, affects^expression
21449848	benazepril (MESH:C044946)	AGTR1 (185)	[AGT gene polymorphism co-treated with ACE2 gene polymorphism] affects the susceptibility to benazepril	affects^cotreatment, affects^response to substance

Section 4: Was the drug identified by manual or computation screen?

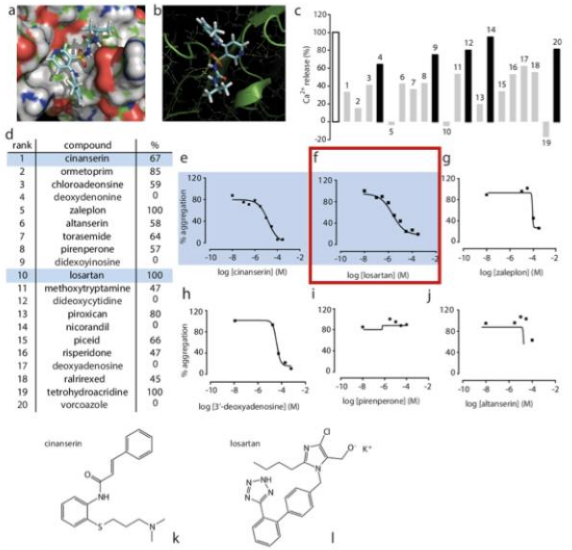
Disease	**COVID-19**
PMID, PMCID	**Evidence Sentences**
32081428 PMC7092824	Among them are captopril, perindopril, ramipril, lisinopril, benazepril, and moexipril. Although these drugs primarily target ACE, a homolog of ACE2 with 42% sequence identity and 61% sequence similarity in the catalytic domain, they may be effective toward ACE2 as well By using a molecular docking approach, an earlier study identified N-(2-aminoethyl)-1 aziridine-ethanamine as a novel ACE2 inhibitor that effectively blocks the SARS-CoV RBD-mediated cell fusion. This has provided a potential candidate and lead compound for further therapeutic drug development. Meanwhile, biochemical and cell-based assays can be established to screen chemical compound libraries to identify novel inhibitors.

Disease	**Cardiovascular Disease**
PMID, PMCID	**Evidence Sentences**
22800722 PMC7102827	Currently, there are more than 10 ACE inhibitors marketed that are widely used as first-line therapy for cardiovascular diseases, including hypertension, heart failure, heart attack and left ventricular dysfunction. According to the functional moiety, they are divided into three types: thiol (captopril), carboxylate (benazepril, enalapril, lisinopril, moexipril, perindopril, quinapril, ramipril, trandolapril) or phosphate (fosinopril). Structure-based drug screening has identified two ACE2 activators: a xanthenone (1-[(2-diethylamino)ethyl-amino]-4-(hydroxymethyl)-7-[(4-methylphenyl)sulphonyloxy]-9H-xanthene-9-one; XNT) and resorcinolnaphthalein.XNT hydrogen bonds with ACE2 residues Lys94, Tyr196, Gly205 and His 195, and resorcinolnaphthalein is involved in three hydrogen bonds with residues Gln98, Gln101 and Gly205. XNT and resorcinolnaphthalein modulate ACE2 activity possibly by two mechanisms.

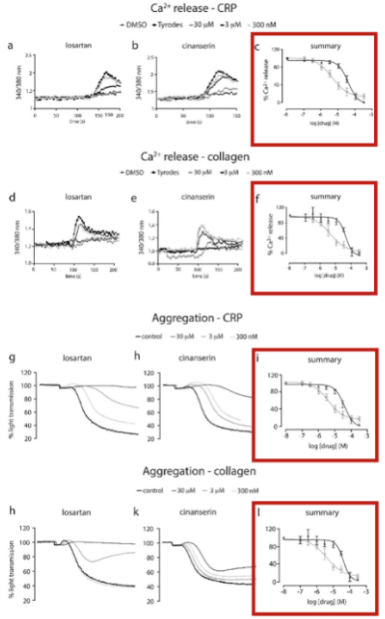
Section 5: Who is studying the drug? (Source/lab name)

Researcher	Affiliation
Robert L Kruse	Division of Transfusion Medicine, Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD.
Hua-Hao Fan	Beijing Advanced Innovation Center for Soft Matter Science and Engineering, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China.
Li-Qin Wang	Gansu Provincial Center for Disease Control and Prevention, Lanzhou, Gansu, China.
Wen-Li Liu	1Department of Dermatology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, 710061 Shaanxi China.
Xiao-Ping An	College of Animal Science, Inner Mongolia Agricultural University, Hohhot, 010018, China.
Zhen-Dong Liu	Key Laboratory of Saline-alkali Vegetation Ecology Restoration, Ministry of Education/Alkali Soil Natural Environmental Science Center, Northeast Forestry University, Harbin 150040, China. liu304418091@126.com.
Xiao-Qi He	
Li-Hua Song	School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China. lihuas@sjtu.edu.cn.
Yi-Gang Tong	Beijing Advanced Innovation Center for Soft Matter Science and Engineering (BAIC-SM), College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, P. R. China.
Pei-Fang Wei	
Jin Soo Shin	Infectious Diseases Therapeutic Research Center, Korea Research Institute of Chemical Technology (KRICT), Daejeon 34114, Korea.
Eunhye Jung	Virus Research Group, Korea Research Institute of Chemical Technology, Daejeon, Republic of Korea.
Meehyein Kim	Infectious Diseases Therapeutic Research Center, Korea Research Institute of Chemical Technology (KRICT), Daejeon 34114, Korea.
Ralph S Baric	Department of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill NC 27599, USA.
Yun Young Go	Virus Research Group, Korea Research Institute of Chemical Technology, Daejeon, Republic of Korea. yygo@krict.re.kr.
Junwen Luan	Institute of Basic Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250062, Shandong, China.
Yue Lu	Department of Epigenetics and Molecular Carcinogenesis, University of Texas M.D. Anderson Cancer Center, Science Park, Smithville, TX, 78957, USA.
Xiaolu Jin	Institute of Basic Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China.
Leiliang Zhang	Institute of Basic Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250062, Shandong, China. Electronic address: armzhang@hotmail.com.
Renhong Yan	Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China.
Yuanyuan Zhang	College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan, China.
Yaning Li	State Key Laboratory of Satellite Navigation System and Equipment Technology, Shijiazhuang 050081, China.
Lu Xia	Shanghai Public Health Clinical Center, Fudan University, Shanghai, People's Republic of China.
Yingying Guo	Department of Cardiothoracic Surgery, The Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, China.
Qiang Zhou	Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.

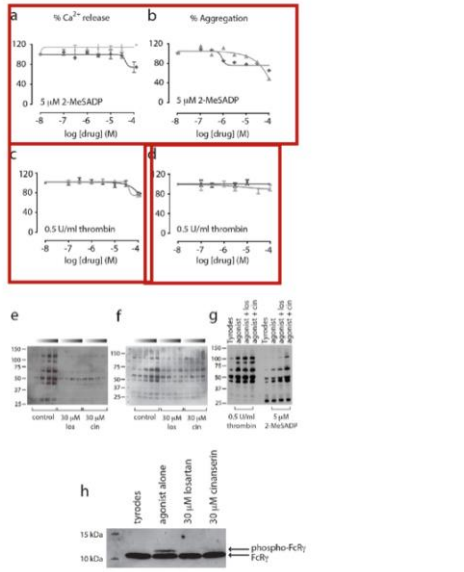
Section 6: In vitro Data available (cell line used, assays run, viral strain used, cytopathic effects, toxicity, LD50, dosage response curve, etc)



Src: PMC4074120 Fig. 1. In silico identifies GPVI antagonist. Representative image capture of in silico docking into GPVI using Glide, with space filling model is shown in a, and H-bonding to relevant side chains is detailed in b. The 20 highest ranking compounds were screened for effects on Ca2+ release by the GPVI-specific agonist CRP-XL (10 mg/ml) (c and d, % refers to percent inhibition of Ca2+ release). Maximum Ca2+ release is shown in white, compounds that inhibited Ca2+ release by , 50% or more are in grey, and the remainder in black. Commercially available compounds that inhibited CRP-XL-induced Ca2+ release .50% were further screened by light transmission aggregometry to identify compounds displaying dose-dependent inhibition (e-j). Examples are shown of weak antagonism (g and h) and false positives (i and j). Cinanserin (l) and losartan (k) were taken on for further study.



Src: PMC4074120 Fig. 2. Losartan and cinanserin inhibit GPVI-mediated cell activation. Washed human platelets were loaded with fura2-AM and screened for drug-mediated inhibition of Ca2+ release by 1 mg/ml CRP-XL (n = 3, 6 SEM, representative traces and summary, a–c) and 1 mg/ml collagen (n = 3, SEM, representative traces and summary, d–f); losartan (&) and cinanserin (m). To measure aggregation, washed human platelets were incubated with drug for one minute prior to the addition of 1 mg/ml CRP-XL (representative traces and summary shown in g–i) or 1 mg/ml collagen (representative traces and summary shown in j–l).



Src: PMC4074120 Fig. 3. Losartan and cinanserin demonstrate selectivity for GPVI. Ca2+ release and aggregations were carried out with 5 mM of the P2Y12 receptor agonist 2-MeSADP (a, Ca2+ release and b, aggregation), or 0.5 U/ml of the PAR1 and PAR4 receptor agonist thrombin (c, Ca2+ release and d, aggregation). Losartan (X); cinanserin (m), n = 3, 6 SEM. For global tyrosine phosphorylation, washed human platelets were incubated with drug or vehicle alone before addition of 1 mg/ml CRP-XL or collagen. Samples were collected at 10, 30, 60 or 90 seconds (as indicated by the graduated bars with time increasing to the right) in ice cold 26 lysis buffer and separated on 4–12% NuPage gels under reducing conditions. Tyrosine phosphorylation was visualized with 4G10 anti-phosphotyrosine antibody. Losartan and cinanserin reduce CRP-XL- (e) and collagen-(f) induced global tyrosine phosphorylation, but have no effect on thrombin or 2-MeSADP induced global tyrosine phosphorylation (g). Both drugs reduce FcRc phosphorylation (h), (unphosphorylated, lower band; phosphorylated, upper band).

Disease	**Cardiovascular disease**
PMID, PMCID	Evidence Sentences
22800722 PMC7102827	The in vitro half-maximal inhibitory concentration (IC50) values of food-derived ACE inhibitory peptides are about 1000-fold higher than that of synthetic captopril but they have higher in vivo activities than would be expected from their in vitro activities..... Germinal ACE depends on chloride to a lesser extent compared with the C domain of sACE. Cushman and Cheung reported an optimal in vitro ACE activity of rabbit rung acetone extract in the presence of 300 mM NaCl at pH 8.1-8.3…

Section 7: Animal Data Available (what animal model, LD50, dosage response curve, etc)

Disease	**Feline Infectious Peritonitis**
PMID, PMCID	Evidence Sentences
21216644 PMC7129202	The in vitro half-maximal inhibitory concentration (IC50) values of food-derived ACE inhibitory peptides are about 1000-fold higher than that of synthetic captopril but they have higher in vivo activities than would be expected from their in vitro activities..... Germinal ACE depends on chloride to a lesser extent compared with the C domain of sACE. Cushman and Cheung reported an optimal in vitro ACE activity of rabbit rung acetone extract in the presence of 300 mM NaCl at pH 8.1-8.3…

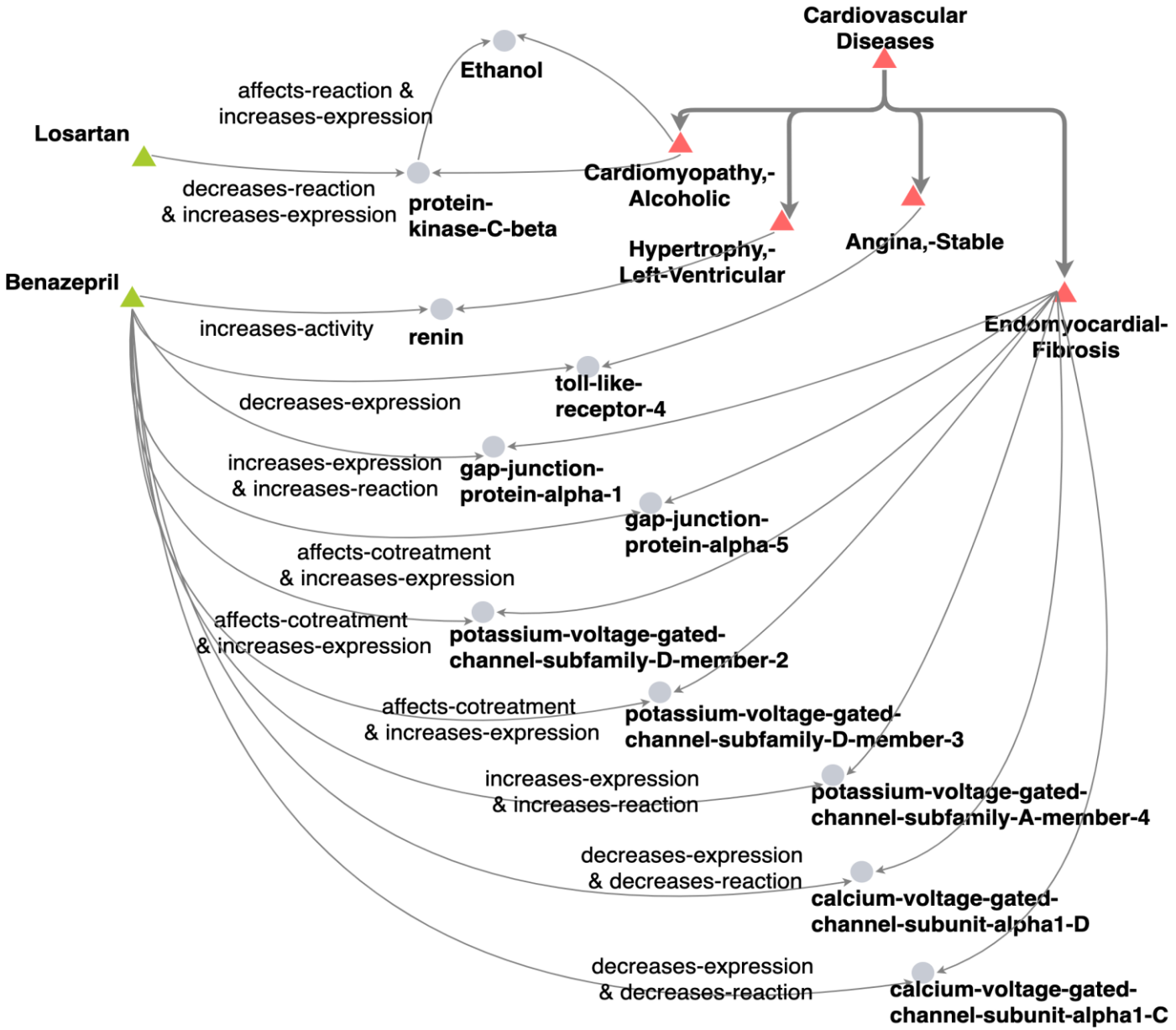
Section 8: Clinical trials on going (what phase, facility, target population, dosing, intervention etc)

Disease	**Cardiovascular disease**
PMID, PMCID	Evidence Sentences
18804122 PMC7112668	On the basis of several large-scale clinical trials, it is now accepted that chronic inhibition of the RAS can provide neuroprotection, with reduced occurrence of stroke in high-risk populations.

Disease	**COVID-19**
PMID, PMCID	Evidence Sentences
32336612 PMC7167588	Two trials of losartan as additional treatment for SARS-CoV-2 infection in hospitalized (NCT04312009) or not hospitalized (NCT04311177) patients have been announced, supported by the background of the huge adverse impact of the ACE Angiotensin II AT1 receptor axis over-activity in these patients.
32350632 PMC7189178	To address the role of angiotensin in lung injury, there is an ongoing clinical trial to examine whether losartan treatment affects outcomes in COVID-19 associated ARDS (NCT04312009).
32439915 PMC7242178	Losartan was also the molecule chosen in two trials recently started in the United States by the University of Minnesota to treat patients with COVID-19 (clinical trials.gov NCT04311177 and NCT 104312009).

Disease	**Cardiovascular Disease**
PMID, PMCID	Evidence Sentences
32034644 PMC7171054	At least 17 warnings to date have been listed on the Food and Drug Administration recall website (https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts). ARBs, such as valsartan and losartan, represent a class of medications that in randomized controlled clinical trials (RCTs) have been shown to reduce blood pressure (BP) in hypertensive patients and impart cardiovascular benefits in diabetic nephropathy, systolic heart failure, left ventricular dysfunction, and following stroke.

Section 9: Has the drug shown evidence of systemic toxicity?



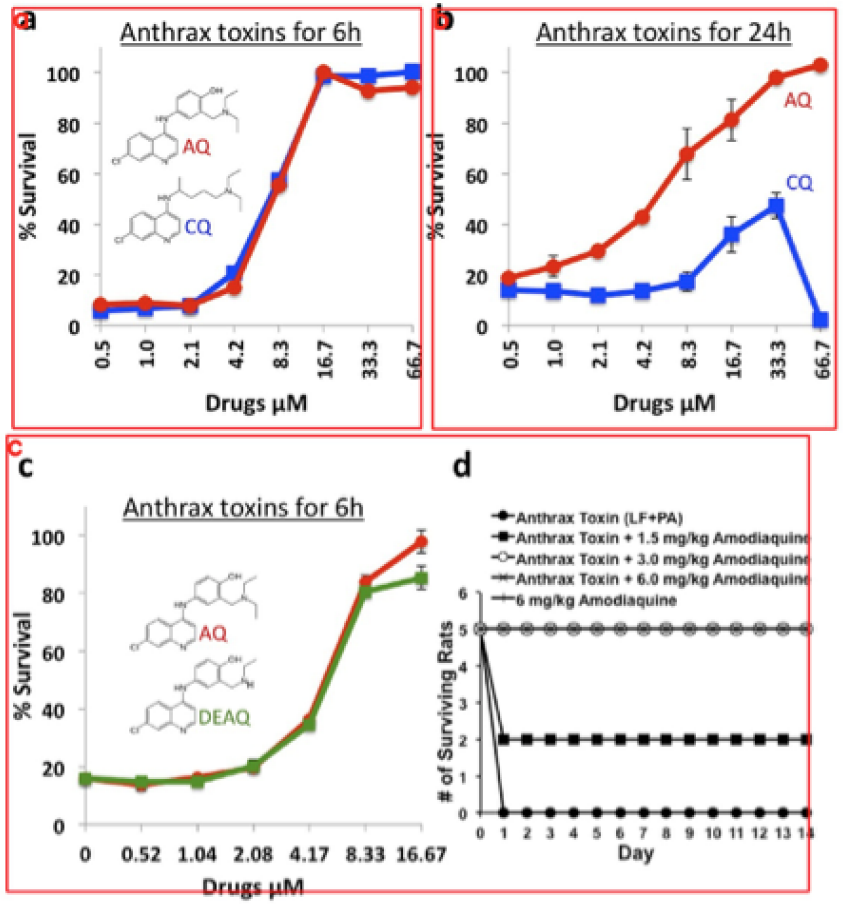
Section 9: Has the drug shown evidence of systemic toxicity?

Disease	Cardiovascular Disease
PMID, PMCID	Evidence Sentences
11883791	The effects of chronic treatment with losartan, an angiotensin II type 1 (AT1) receptor antagonist, and benazepril, an angiotensin converting enzyme (ACE) inhibitor, on target-organ damage and abnormal circadian blood pressure (BP) rhythm were compared in stroke-prone spontaneously hypertensive rats (SHRSP).

Disease	Cardiovascular Disease
PMID, PMCID	Evidence Sentences
11883791	The effects of chronic treatment with losartan, an angiotensin II type 1 (AT1) receptor antagonist, and benazepril, an

Disease	Cardiovascular Disease
PMID, PMCID	Evidence Sentences
11883791	The effects of chronic treatment with losartan, an angiotensin II type 1 (AT1) receptor antagonist, and benazepril, an angiotensin converting enzyme (ACE) inhibitor, on target-organ damage and abnormal circadian blood pressure (BP) rhythm were compared in stroke-prone spontaneously hypertensive rats (SHRSP).

Section 9: Has the drug shown evidence of systemic toxicity?



Src: PMC4550849 Fig 2. The anti-LF-PA efficacy of Amodiaquine (AQ) and its metabolite Desethylamodiaquine (DEAQ) in intro and in vivo. Amodiaquine and JHCCL available structural analogs were tested for their ability to inhibit LF-PA-mediated cytotoxicity. RAW264.7 cells were seeded at 1×104 cells/well on 96-well plates and the following day were incubated with indicated doses of Amodiaquine or Chloroquine (CQ) for 1h, followed by either 6h (a) or 24h (b) intoxication with 500ng/mL PA+500ng/mL LF. Cell viability was determined by MTT assay (Materials and Methods) and is shown as the percentage of survivors relative to cells not treated with drugs. AQ and DEAQ were tested for their ability to inhibit LF-PA-mediated toxicity in (c) RAW264.7 cells and in (d) Sprague-Dawley rats. RAW264.7 cells were seeded at 1 × 104 cells/well on 96-well plates and the following day were incubated with indicated doses of AQ or DEAQ for 1 h, followed by 6h intoxication with 500ng/mL PA+500ng/mL LF. Survival curves of groups of male Sprague-Dawley rats challenged intravenously with 12µg LF and 40µg PA along with varying amounts of AQ (1.5, 3.0, and 6.0mg/kg). Control groups of rats either received toxin only or 6.0mg/kg of AQ only.

Section 10: Funding source

PMCID	Trial Type	Funding source
PMC4074120	Vitro Data	BHF Centre for Research Excellence (Oxford)
PMC7101871	Vitro Data	Hong Kong General Research Fund HKU 768910M, L & T Charitable Foundation and the House of INDOCAFE
PMC4762923	Vitro Data	Major Research Plan of the National Natural Science Foundation of China (No. 91439207) and National Natural Science Foundation of China (Nos. 81300246, 81400314, 81270331, 31370931, 81300196, 81470494)
PMC7167588	Animal Data	Study supported in part by the no-profit Fondazione Umbra Cuore e Ipertensione-ONLUS, Perugia, Italy
PMC7236830	Animal Data	National Heart, Lung, and Blood Institute
PMC7088148	Animal Data	This work was supported by a grant from the Fellowship Research Fund of Sparrow Hospital, Lansing, MI (to C.G).
PMC7167588	Clinical trials	Study supported in part by the no-profit Fondazione Umbra Cuore e Ipertensione-ONLUS, Perugia, Italy
PMC7189178	Clinical trials	
PMC7242178	Clinical trials	
PMC7171054	Clinical trials	

Section 11: List of relevant sources to pull data from

1. https://www.ncbi.nlm.nih.gov/research/pubtator/ (Wei, C. H., Allot, A., Leaman, R., & Lu, Z. (2019). PubTator central: automated concept annotation for biomedical full text articles. Nucleic acids research, 47(W1), W587-W593.)
2. https://www.semanticscholar.org/cord19 (Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Burdick, D., … Kohlmeier, S. (2020). CORD-19: The COVID-19 Open Research Dataset. In Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020. Online: Association for Computational Linguistics. https://www.aclweb.org/anthology/2020.nlpcovid19-acl.1)

Ethical Considerations

- Required Workflow for Using Our System
  - Our knowledge discovery tool provides investigative leads, not final results for clinical use. COVID-KG (and all knowledge discovery tools for biomedical applications) is not meant to be used for direct clinical applications on any human subjects.
  - Our tool provides source and rich evidence sentences for each node and link in the KG. To curtail potential harms caused by extraction errors, users of the knowledge graphs should double-check the source information and verify the correctness of the discovered leads before launching expensive experimental studies.
- Limitations of System Performance and Data Collection
  - Our system can effectively convert a large number of scientific papers into knowledge graphs, and scale as literature volume increases. However, none of our extraction components is perfect, they produce about 6%-22% false alarms and misses as reported in section 2.
  - Proper use of the technology requires that input documents are legally and ethically obtained. The input data to our system is peer-reviewed publicly available scientific articles. An additional potential harm could come from the output of the system being used in ways that magnify the system errors or bias in its training data. Our system output is intended for human interpretation. Incorporating the system's output into an automatic decision-making system without human validation could be harmful.
  - The performance of our system components as reported is based on the specific benchmark datasets, which could be affected by such data biases. A general approach to ensure proper application should: incorporate ethics considerations as the first-order principles in every step of the system design, maintain a high degree of transparency and interpretability of data, algorithms, models, and functionality throughout the system, make software available as open source for public verification and auditing, and explore countermeasures to protect vulnerable groups.
  - In addition, our system output may include some biases from the sources, namely some papers may have got published due to the biases from peer reviewers. We plan to extend our framework to include fact-checking to enable practitioners and researchers to access up-to-the-minute information.
  - Finally, the queries (i.e., the lists of candidate drugs and proteins/genes) are provided by the users who might have a bias in their selection. We require the users to carefully examine source information (author, publication date, etc.) and detailed evidence (contextual sentences and documents) associated with the extracted connections.

Thank you!

Demo video: http://159.89.180.81/demo/covid/Covid-KG_DemoVideo.mp4,

Project website: http://blender.cs.illinois.edu/covid19,

Drug Repurposing Report: http://blender.cs.illinois.edu/covid19/DrugRepurposingReport_V2.0.docx