

# A Comprehensive Update on **CIDO**: The Community-based Coronavirus Infectious Disease Ontology

***Yongqun “Oliver” He***

*University of Michigan Medical School*

Ann Arbor, MI 48109 USA



# Acknowledgements

## Collaborators & co-authors:

### *University of Michigan:*

- Yongqun “Oliver” He\*
- Anthony Huffman
- Medical School)
- Yingtong Liu
- Edison Ong
- Yang Wang (also Guizhou
- Philip Huang
- Long Tran
- Jinyang Du
- Zalan Shah
- Easheta Shah
- Roshan Desai
- Hsin-Hui Huang
- Gil Omenn
- Brian Athey

- Hong Yu\*, Yang Wang, Xianwei Ye (Guizhou Medical School)
- Asiyah Yu Lin (National Center for Ontological Research)
- Darren A. Natale (Georgetown University Medical Center)
- John Beverley, Eric Merrell, Barry Smith (Univ. at Buffalo)
- Ling Zheng, Yehoshua Perl (New Jersey Institute of Technology)
- Zhigang Wang, Xiaolin Yang (Chinese Academy of Med. Sci.)
- Yujia Tian (Rutgers University)
- William D. Duncan (University of Florida)
- Sivaram Arabandi (OntoPro LLC)
- Lynn M. Schriml (Univ. of Maryland School of Medicine)
- Jie Zheng (UPenn)
- Anna Maria Masci (National Inst. of Environmental Health Sci.)
- Liwei Wang, Hongfang Liu (Mayo Clinic)
- Fatima Z. Smaili, Robert Hoehndorf (King Abdullah Univ.)
- Zoë May Pendlington, Paola Roncaglia (EBI)
- Jiangan Xie (Chongqing Univ. of Posts and Telecommun.)
- Yi-Wei Tang (Cepheid, Danaher Diagnostic Platform, Shanghai, China)
- Suyuan Peng, Luxia Zhang (Peking University)
- Luonan Chen (Shanghai Institutes for Biological Sciences)
- Junguk Hur (Univ. of North Dakota)

Paper accepted for publication in  
*Journal of Biomedical Semantics*.

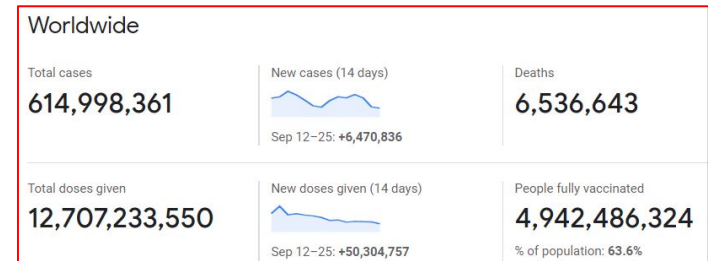
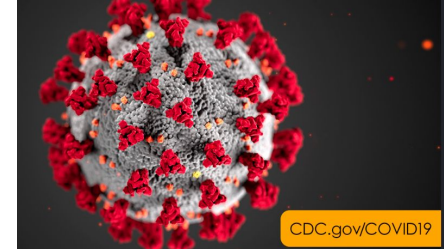
# Acknowledgements

## **Funding:**

This project is supported by NIH grants 1UH2AI132931 (to YH); U24CA210967 and P30ES017885 (to GSO); R01GM080646, 1UL1TR001412, 1U24CA199374, and 1T15LM012495 (to BS); the National Natural Science Foundation of China 61801067 (to JX); the Natural Science Foundation of Chongqing CSTC2018JCYJAX0243 (to JX); the non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences 2019PT320003 (to HY); and Undergraduate Research Opportunity Program (UROP) and University of Michigan Medical School Global Reach award (to YH). The work of ZMP and PR was supported by Open Targets (OTAR005).

# COVID-19: Coronavirus Disease 2019

- Caused by SARS-CoV-2 coronavirus
- First found in Wuhan, China, in December 2019
- COVID-19 Pandemic
  - Declared by WHO on March 11
  - At the time: 118,326 confirmed cases and 4,292 deaths globally
  - As of 9/26/2022: >614 mill. cases, >6.5 mill deaths



**As of 9/26/2022**

# Other Coronavirus Diseases

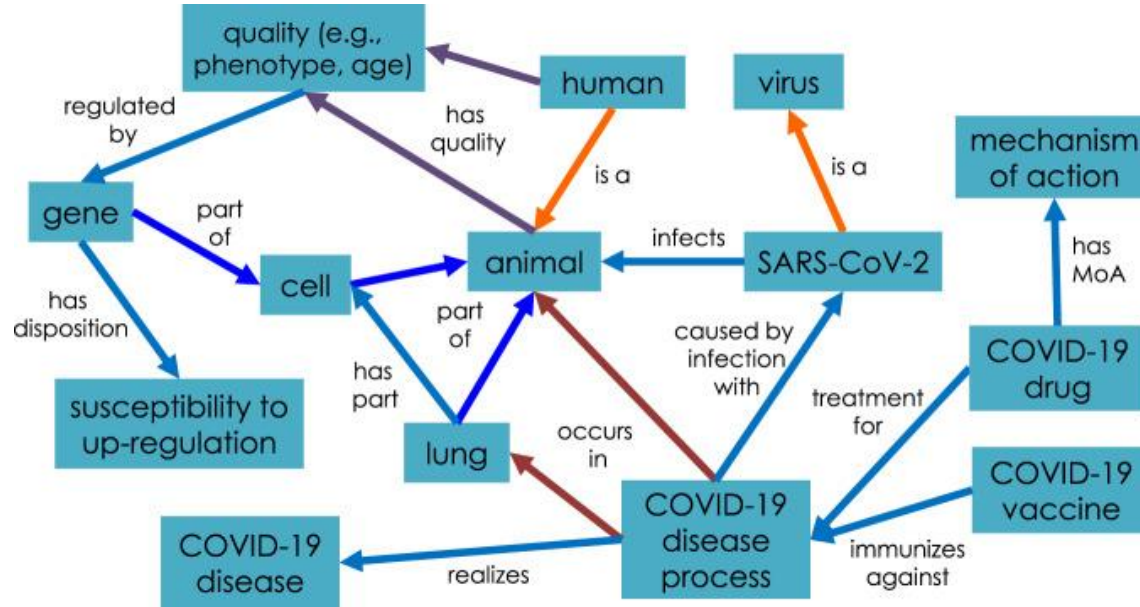
- **SARS:** Severe Acute Respiratory Syndrome
  - Emerged in China in Nov. 2002, lasted 8 months
  - 8,098 confirmed cases in 29 countries, 774 deaths
  - Case-fatality rate: 9.6%
- **MERS:** Middle East Respiratory Syndrome
  - Emerged Saudi Arabia in June 2012
  - 2,260 confirmed cases in 27 countries, 803 deaths
  - Case-fatality rate: 35.5%
- Many other coronaviruses cause mild disease in humans, similar to the common cold

# CIDO: Coronavirus Infectious Disease Ontology

- Community Effort:
  - <https://github.com/CIDO-ontology/cido>
  - An Open Biological and Biomedical Ontology (OBO) Foundry library ontology
- Integrates coronavirus data concerning:
  - Coronaviruses (etiology); Hosts (phenotypes); Reservoirs (Transmission); Host-Coronavirus Interactions;
  - Diagnosis; Drugs; Vaccines;
  - Metadata; Data Standardization

**Ref:** He Y, Yu H, Ong E, Wang Y, Liu Y, Huffman A, Huang H, Beverley J, Hur J, Yang X, Chen L, Omenn GS, Athey B, Smith B. **CIDO**, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. **Scientific Data**. (2020) 7:181.

# CIDO design pattern (2020)



**Ref:** He Y, Yu H, Ong E, Wang Y, Liu Y, Huffman A, Huang H, Beverley J, Hur J, Yang X, Chen L, Omenn GS, Athey B, Smith B. **CIDO**, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Scientific Data*. (2020) 7:181.

# CIDO Statistics:

**Version: 1.0.337 (08-01-2022):**

- [Class](#) (10262)
- [ObjectProperty](#) (370)
- [DatatypeProperty](#) (24)
- [AnnotationProperty](#) (170)
- [Instance](#) (458)

- 11,284 terms in total
  - 1,526 CIDO specific terms
  - Others imported from 53 existing ontologies (note: many terms such as vaccines are from our own or our co-authors' ontologies such as Vaccine Ontology and Protein Ontology.)

Index	Ontology Prefix	Class	ObjectProperty	DatatypeProperty	AnnotationProperty	Instance	Total
1	APOLLO	20	0	0	1	0	21
2	BFO	20	20	0	2	0	67
3	CHEBI	2,009	0	0	0	0	2,009
4	CIDO	1,429	87	0	4	1	1,526
5	CL	0	0	0	0	0	0
6	CMO	1	0	0	0	0	1
7	DOID	100	0	0	0	0	100
8	DRON	122	0	0	1	0	123
9	ENVO	0	0	0	0	0	0
10	GAZ	1	0	0	0	1	2
11	GENO	1	0	0	0	0	1
12	GO	500	0	0	0	0	500
13	HP	246	0	0	0	0	246
14	IAO	21	0	1	10	14	58
15	ICDO	1	0	0	0	0	1
16	ICD	0	0	0	1	0	1
17	IDO	23	0	0	0	0	23
18	IMPATH	2	0	0	0	0	2
19	NCBITaxon	1,790	0	0	0	0	1,790
20	NCIT	2	0	0	0	0	2
21	NCIF-RT-owl	1,010	11	17	17	422	1,477
22	OAE	0	0	0	0	0	0
23	OBCS	11	0	0	0	0	11
24	OBI	200	11	1	0	0	217
25	OBIB	1	0	0	0	0	1
26	OGG	37	2	0	0	0	41
27	OGMS	0	0	0	0	0	0
28	OMABIS	2	0	0	0	0	2
29	OMSE	0	0	0	0	0	0
30	OPM	124	0	0	0	0	124
31	PATO	20	0	0	0	0	20
32	PR	1,848	0	0	0	0	1,848
33	RO	0	103	0	0	0	103
34	RTS	0	0	0	0	0	0
35	SO	0	0	0	0	0	0
36	SYMP	1	0	0	0	0	1
37	TEMP	0	4	0	0	0	4
38	TRANS	2	0	0	0	0	2
39	UBERON	122	0	0	0	0	122
40	UBPROP	0	0	0	0	0	0
41	UO	1	0	0	0	1	2
42	VO	336	11	0	4	16	367
43	Viro	0	0	0	0	0	0
44	chabi	0	2	0	0	0	2
45	core	0	0	0	0	0	0
46	doap	0	0	0	0	1	1
47	doal	1	0	0	0	0	1
48	ncbitaxon	0	0	0	0	1	1
49	obol:Owl	0	1	0	0	0	1
50	owl	1	1	0	0	0	2
51	pr	0	1	0	0	0	1
52	protege	0	0	0	0	0	0
53	rd-schema	0	1	0	0	0	1
54	rxowl	0	1	0	0	0	1
55	taxsim	0	0	0	0	0	0
56	NePrefix	0	1	0	0	0	1
Total	-	10,262	370	24	170	458	11,284

<https://ontobee.org/ontostat/CIDO>



# Architecture of CIDO

Aligns with  
Basic  
Formal  
Ontology  
(BFO)

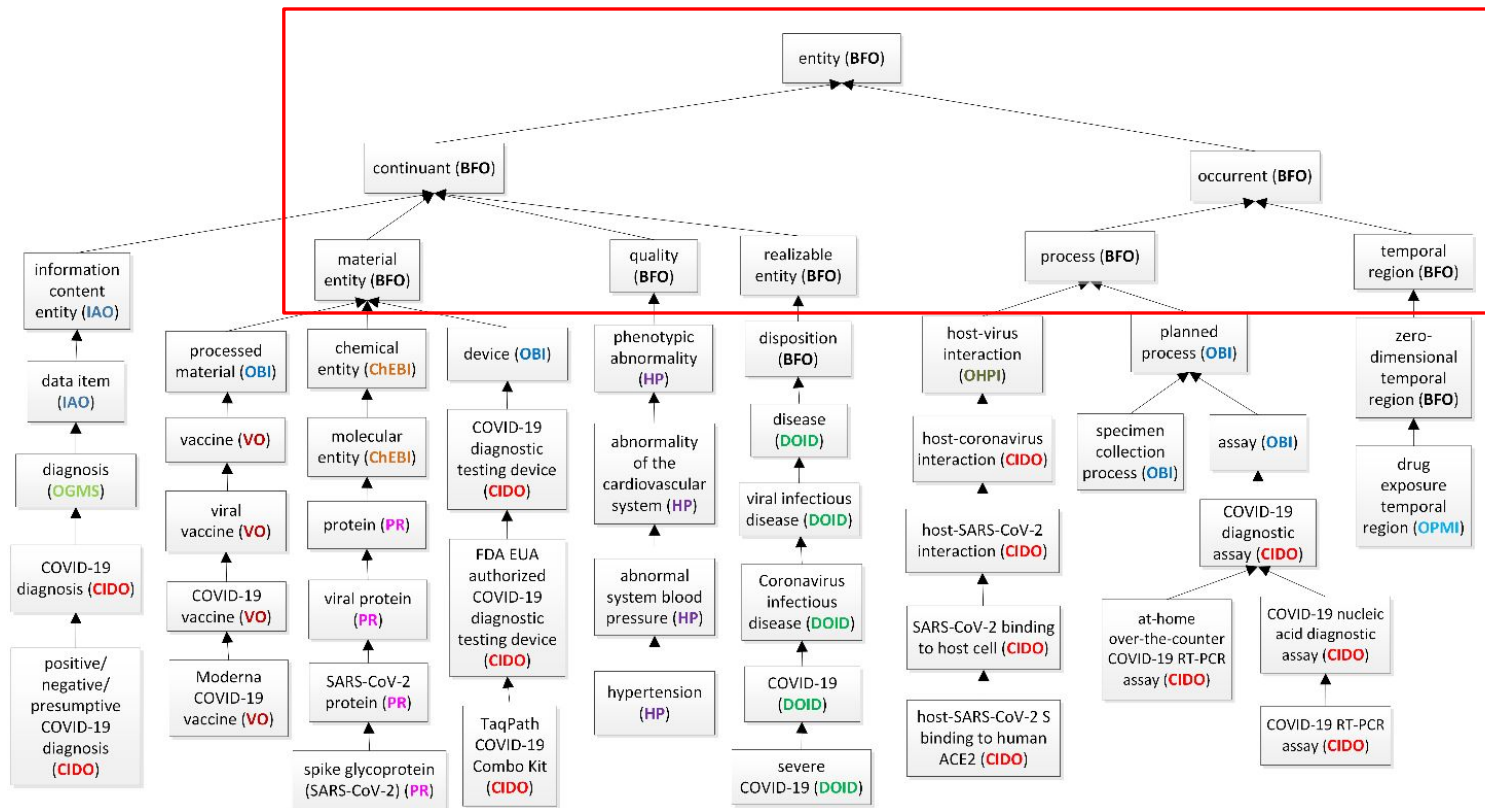
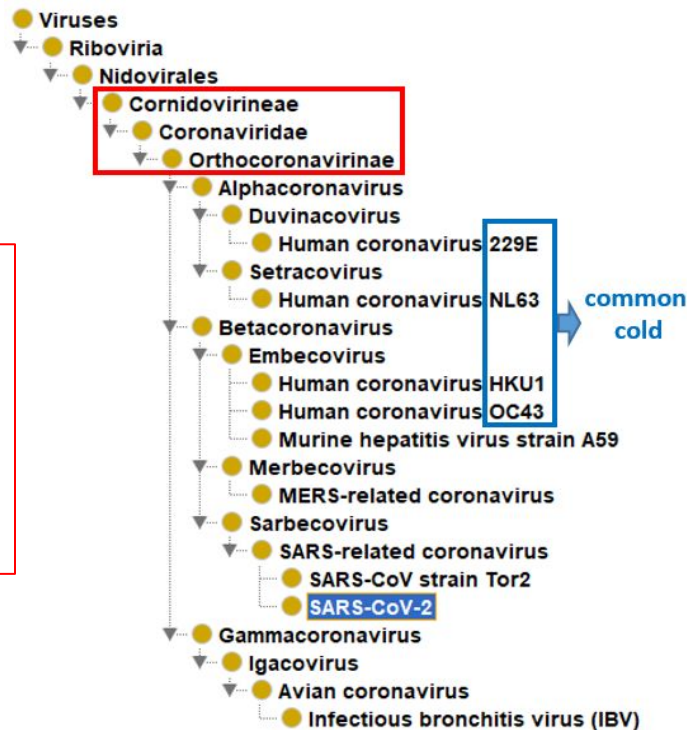


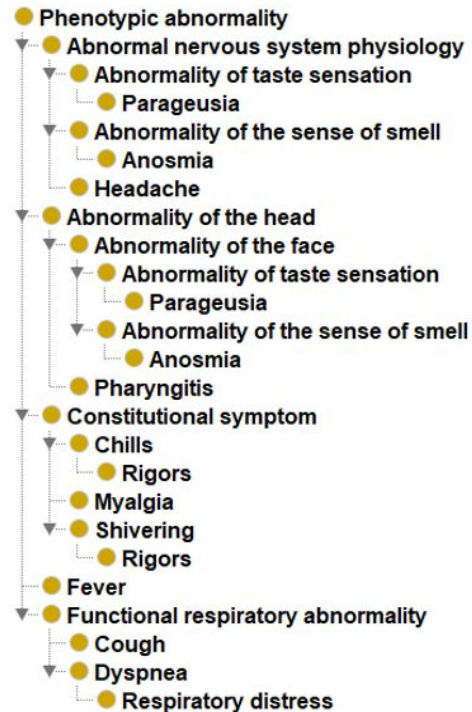
Fig. 1.

# Representing Coronaviruses and COVID-19 Phenotypes



NCBITaxon  
hierarchy  
imported to  
CIDO,  
representing  
coronaviruses

(A)



HP imports  
represent  
phenotypes  
shown in  
COVID-19

(B)

Strategy used: (1) Only import needed, (2) Semantics & additional terms also imported.

# SARS-CoV-2 proteins and genes

By: Darren A.  
Natale, Oliver He

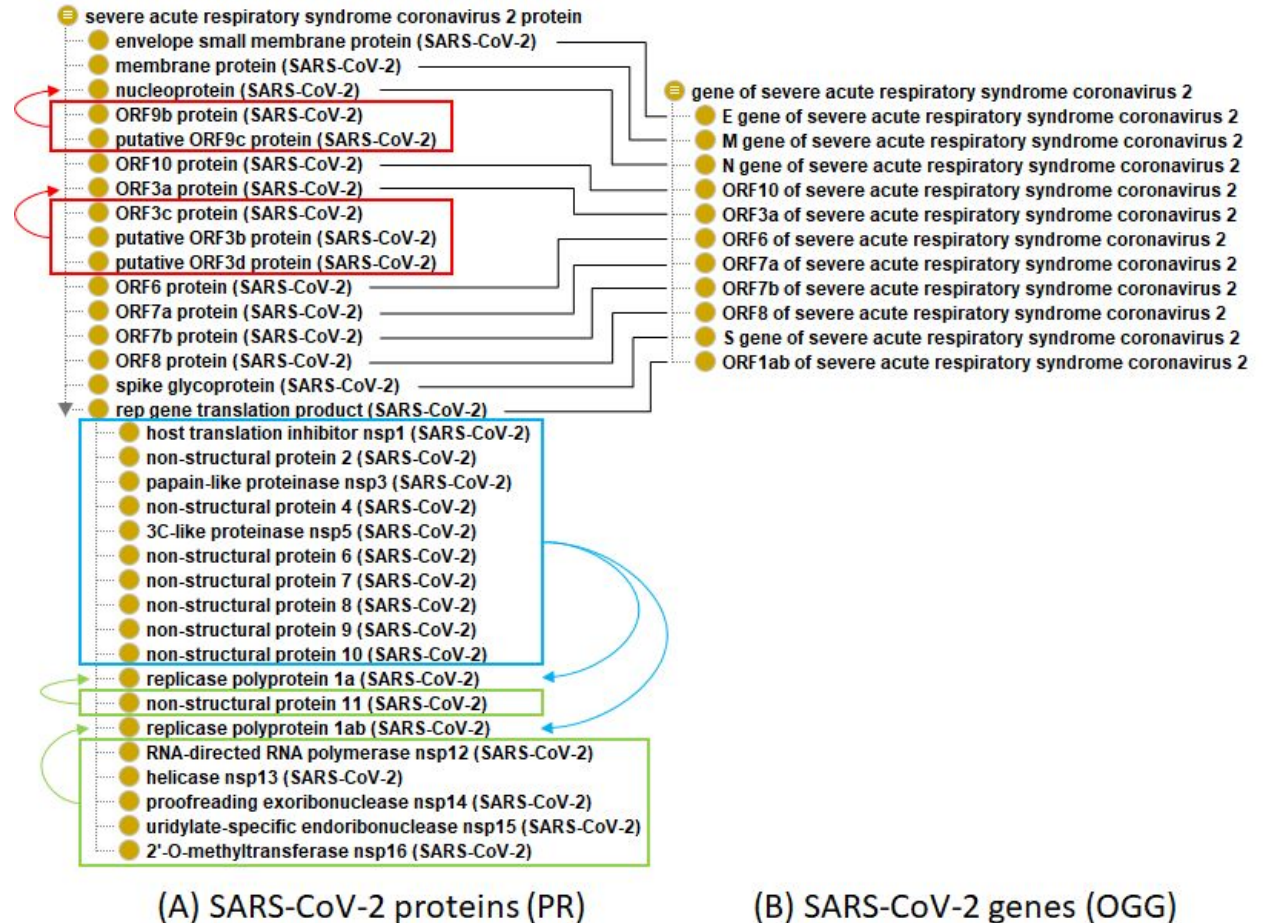


Fig. 2.

# CIDO modeling of AA variants and mutations

## ***‘D-614G in SARS-CoV-2 S protein’:***

- *‘characteristic AA variant of’ some ‘SARS-CoV-2 Omicron variant’*
- *‘is a’ some ‘AA variant in SARS-CoV-2 S protein S1 RBD region’*
- *‘has amino acid position’ value 614*
- *‘has part’ some ‘glycine residue’*
- *‘has mutated from’ some ‘aspartic acid’*

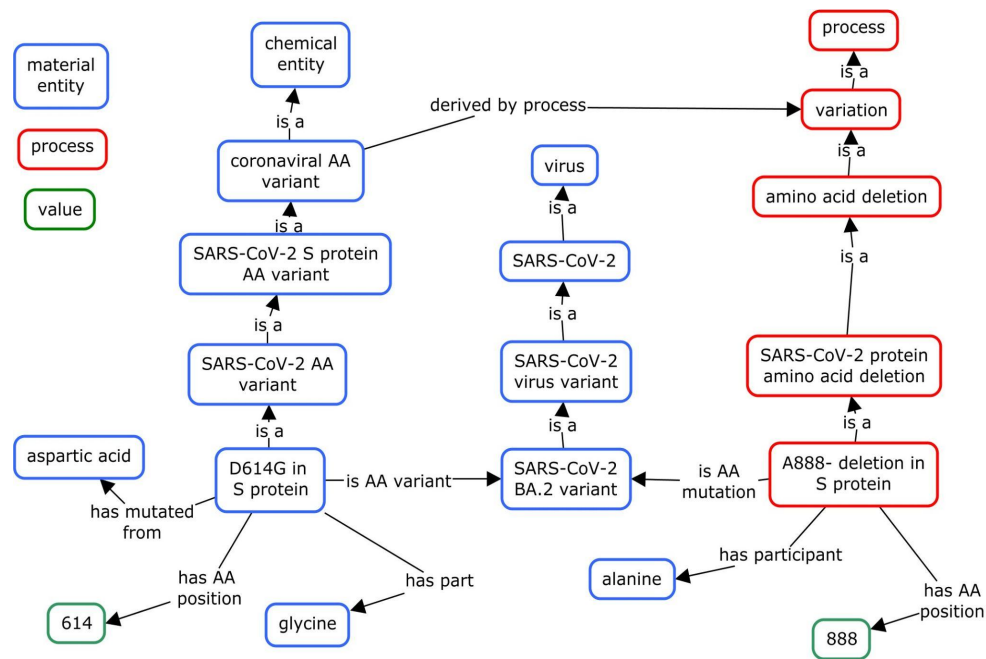


Fig. 3.

# Ontological modeling of epidemiology and public health

*'blood plasma specimen 1' rdf:type 'blood plasma specimen'  
and 'has part' some 'SARS-CoV-2'  
and 'has viral load measurement' value  $10^8$*

By: John  
Beverley, VIDO

Epidemiology:

*'SARS-CoV-2 reference strain: 'has average R0' value 2.69*

*'SARS-CoV-2 Delta variant': 'has average R0' value 5.02*

*'SARS-CoV-2 Omicron BA.1 variant': 'has average R0' value 9.05*

Fig. 3.

# Modeling of COVID-19 diagnostic testing using CIDO

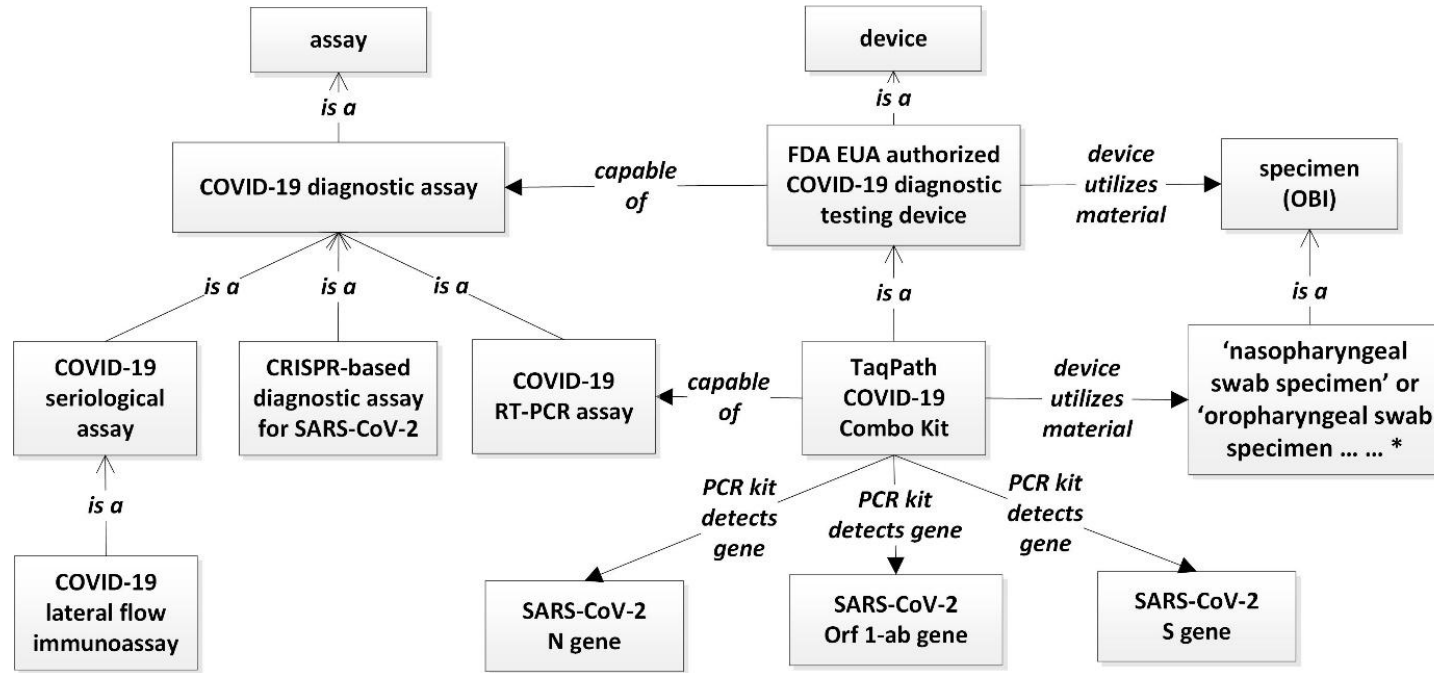


Fig. 4.



# Pfizer/BioNTech vaccine - Comirnaty

## 6. Comirnaty

- Other names: BNT162b2, Tozinameran
- Manufacturer: BioNTech, Pfizer
- First authorized on: December 11, 2020
- Countries where authorized: USA, Canada, and other countries
- Clinical trial ID:
  - Phase IV: [NCT04969250](#)
    - Age subgroups: 18 years and above
    - Location: USA
  - Phase IV: [NCT04952766](#)
    - Age subgroups: 18 years and above
    - Location: France
- Type: mRNA vaccine
- Publications:
  - Borobia *et al.* Immunogenicity and reactogenicity of BNT162b2 boost participants (CombiVacS): a multicentre, open-label, randomised, cc *Lancet*. 2021; 398(10295); 121-130. [PubMed: [34181880](#)].
  - Frencz *et al.* Safety and immunogenicity of the SARS-CoV-2 BNT162b1 younger and older Chinese adults: a randomized, placebo-controlled study. *Nat Med*. 2021; 27(6); 1062-1070. [PubMed: [33888900](#)].
  - Sahin *et al.* COVID-19 vaccine BNT162b1 elicits human antibody at Nature. 2020; 586(7830); 594-599. [PubMed: [32998157](#)].
  - Mulligan *et al.* Phase 1/2 study of COVID-19 RNA vaccine BNT162b1. *Nature*. 2020; 586(7830); 589-593. [PubMed: [32785213](#)].
- Additional links:
  - [FDA Briefing Document](#)
  - [VIOLIN vaccine ID](#)
  - [VO: VO\\_0004987](#) **Vaccine Ontology ID**

### Comirnaty

#### Vaccine Information

- Vaccine Name:** Comirnaty
- Target Pathogen:** SARS-CoV-2
- Target Disease:** COVID-19
- Product Name:** BNT162b2
- Manufacturer:** Pfizer, BioNTech
- Vaccine Ontology ID:** [VO\\_0004987](#)
- CDC CVX code:** 208
- CDC CVX description:** SARS-COV-2 (COVID-19) vaccine, mRNA, spike protein, LNP, preservative free, 30 mcg/0.3mL dose
- Type:** mRNA vaccine
- Status:** Clinical trial
- Host Species for Licensed Use:** Human
- Antigen:** trimerized SARS-CoV-2 receptor-binding domain from S
- Immunization Route:** Intramuscular injection (i.m.)
- Storage:** -70°C ±10°C
- Description:** A SARS-CoV-2 RNA vaccine formed from a lipid nanoparticle-formulated domain

#### Host Response

#### Macaque Response

- Vaccination Protocol:** Groups of six male, 2-4 year old rhesus macaques were immunized on Days 0 and 21. (Vogel *et al.*, 2020)
- Immune Response:** Seven days after Dose 2 (Day 28), the GMCs of S1-binding IgG and 34,668 U/mL (100 µg dose level). Fifty percent virus neutralisation GMTs, measured on Day 28 after Dose 2.

[http://www.violinet.org/vaxquery/vaccine\\_detail.php?c\\_vaccine\\_id=5784](http://www.violinet.org/vaxquery/vaccine_detail.php?c_vaccine_id=5784)

### Vaccine Ontology

Keywords:  Search terms

Class: Comirnaty

Term IRI: [http://purl.obolibrary.org/obo/VO\\_0004987](http://purl.obolibrary.org/obo/VO_0004987)

Definition: A SARS-CoV-2 RNA vaccine formed from a lipid nanoparticle-formulated trimeric domain

Annotations

- definition editor: Philip Huang, Rohit Goru, Oliver He
- alternative term: Tozinameran; BNT162b2; PF-07302048
- definition source: <https://en.wikipedia.org/wiki/Tozinameran>; <https://www.fda.gov/media/14424/nCoV/vaccines/different-vaccines/Pfizer-BioNTech.html>; <http://www.violinet.org/cov19vaxkb/authorized-response/coronavirus-disease-2019-covid-19/pfizer-biotech-covid-19-vaccine>; PubMed: 33888900

Class Hierarchy

```
graph TD
    Thing --> entity
    entity --> continuant
    continuant --> independent_continuant
    independent_continuant --> material_entity
    material_entity --> processed_material
    processed_material --> vaccine
    vaccine --> viral_vaccine
    viral_vaccine --> coronavirus_vaccine
    coronavirus_vaccine --> COVID_19_vaccine
    COVID_19_vaccine --> authorized_COVID_19_vaccine
    authorized_COVID_19_vaccine --> authorized_COVID_19_RNA_vaccine
    authorized_COVID_19_RNA_vaccine --> Moderna_COVID_19_vaccine
    Moderna_COVID_19_vaccine --> Comirnaty
```

<http://www.violinet.org/cov19vaxkb/authorizedcov19vax.php#Comirnaty>

[http://purl.obolibrary.org/obo/VO\\_0004987](http://purl.obolibrary.org/obo/VO_0004987)

# Representative clinical metadata types covered in CIDO

Metadata types	Metadata Examples
person (NCBITaxon_9606)	person ID (OPMI_0000470), gender (PATO_0001894), year of birth (OPMI_0000473), race (NCIT_C17049 ), ethnicity (NCIT_C16564 ), care site (OPMI_0000479), geographic location (GAZ_00000448)
specimen (OBI_0100051)	specimen ID (OBI_0001616), date of specimen collection (OBIB_0000714), anatomical structure (UBERON_0000061)
visit occurrence (OPMI_0000482)	visit occurrence identifier (OPMI_0000483), visit start date (OPMI_0000487), visit end date (OPMI_0000488), preceding visit occurrence (OPMI_0000492), ER visit (OPMI_0000486)
procedure occurrence (OPMI_0000505)	procedure (NCIT_C25218), procedure start date (OPMI_0000508), procedure end date (OPMI_0000510), care provider (OPMI_0000163)
drug exposure (OPMI_0000572) and device exposure (OPMI_0000554)	drug (CIDO_0000167), drug exposure start time (OPMI_0000565), drug exposure end time (OPMI_0000567), medical device (NCIT_C16830), diagnostic kit (CIDO_0000453)
clinical measurement (CMO_0000000)	clinical measurement identifier (OPMI_0000582), care provider (OPMI_0000163), measurement time (OPMI_0000579), measurement unit label (IAO_0000003), measurement date (OPMI_0000580)
observation period (OPMI_0000575)	observation period start date (OPMI_0000577), observation period end date (OPMI_0000578), provenance of observation record (OPMI_0000522)

Table 1



# Host-coronavirus protein-protein interactions (PPIs) and drugs targeting the viral or host proteins

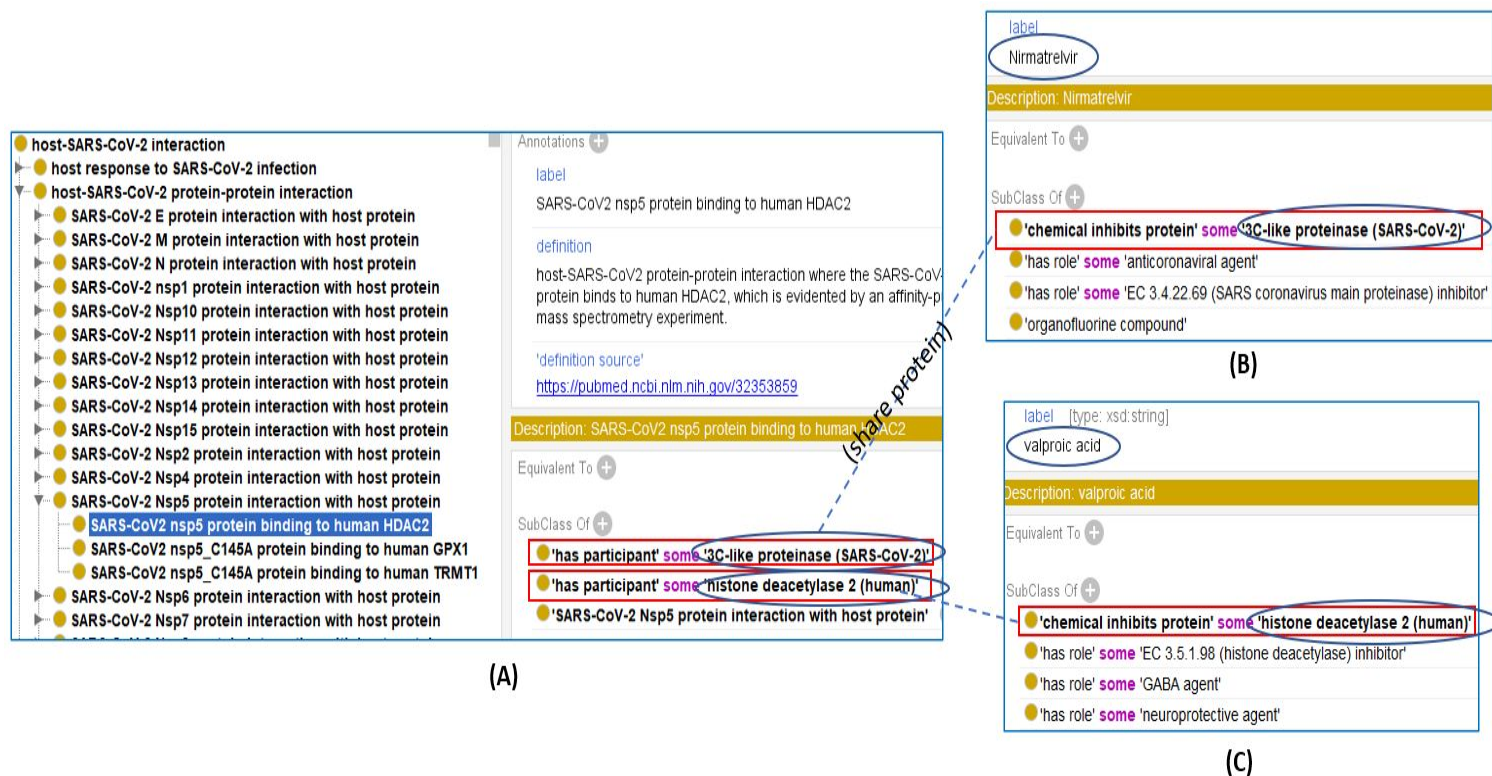


Fig. 5.

# Modeling of COVID-19 diagnostic testing using CIDO

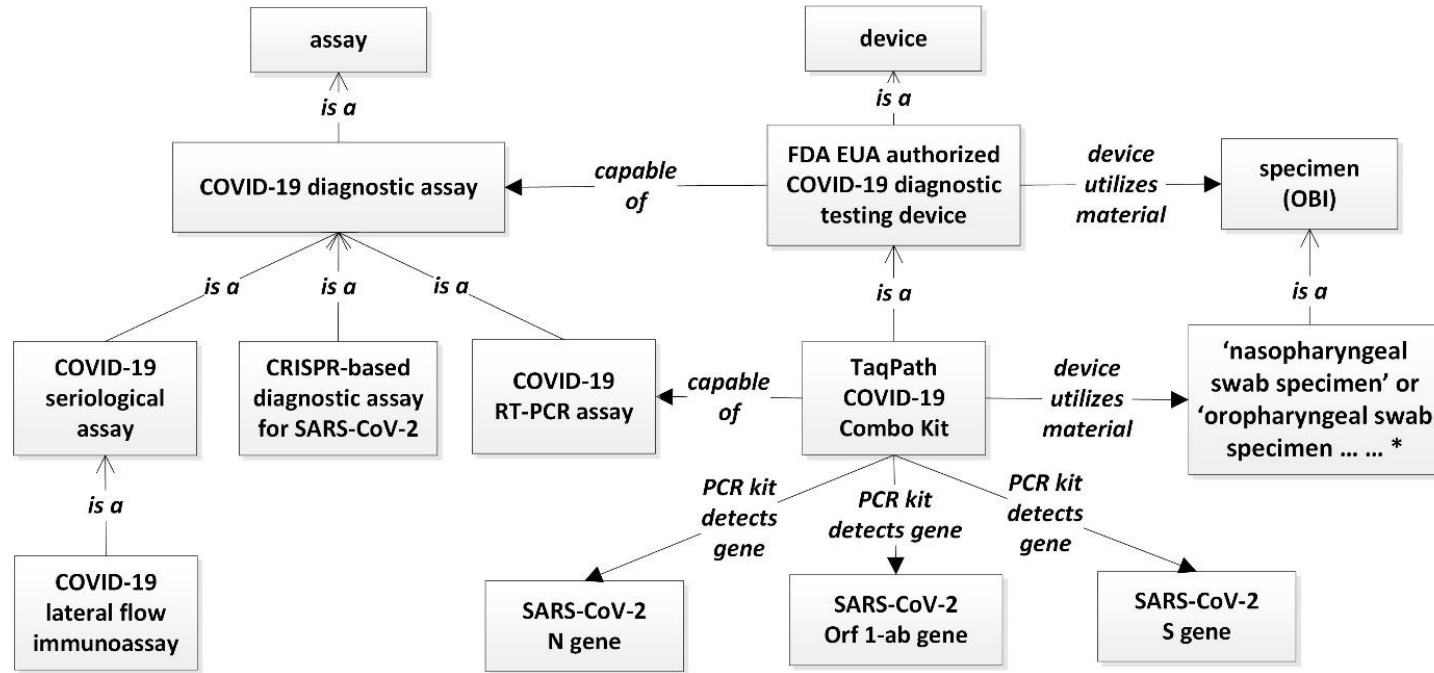


Fig. 4.

# The weighted aggregate taxonomy (WAT) for CIDO (version 1.0.306) with 10,853 concepts

By: Ling Zheng,  
Yehoshua Perl

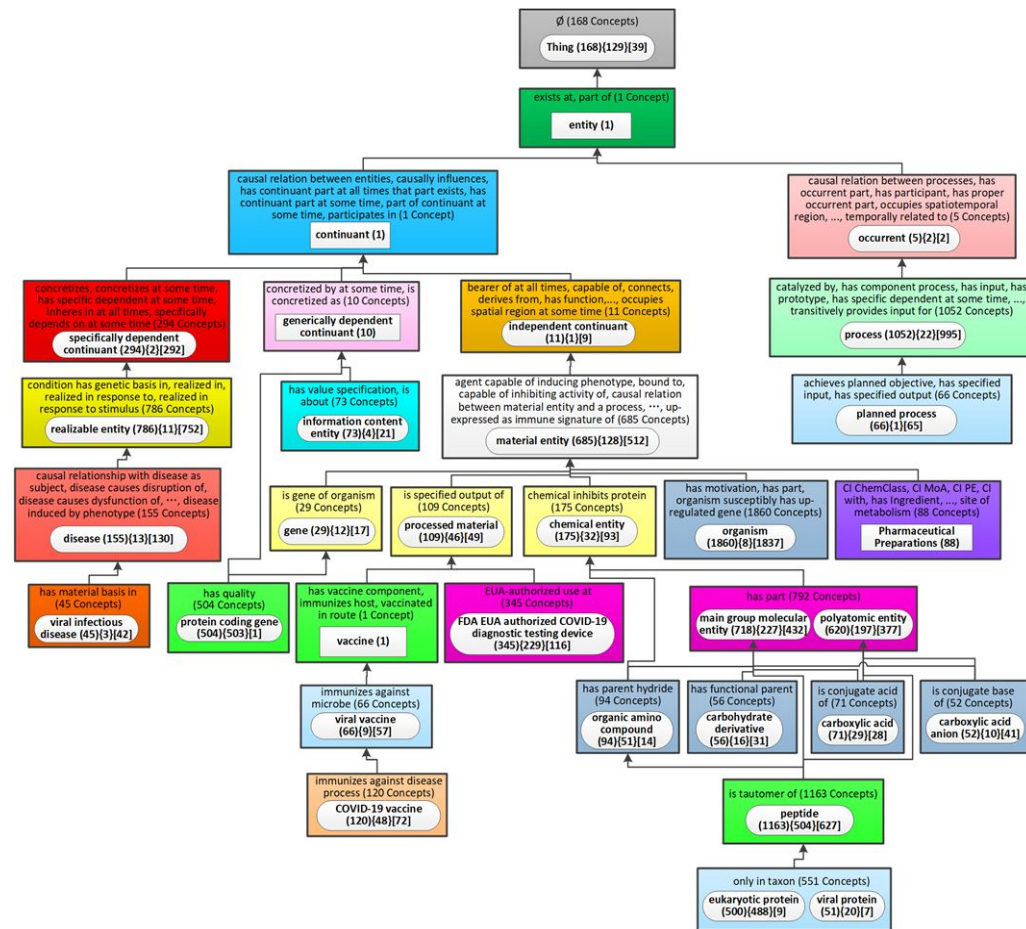


Fig. 6.

# CIDO Use Cases

1. Ontology-based coronavirus-related knowledge and data standardization, annotation, mapping, integration, and inferencing, supporting advanced COVID-19 data analysis
2. CIDO queries for Delta and Omicron differences for better mechanistic understanding of virulence and transmission
3. CIDO-supported NLP for clinical and basic mechanism research
4. CIDO-based machine learning and drug cocktail design for COVID-19 treatment

# DL Query Demos

DL query:

**Query (class expression)**

((is AA mutation of **some** 'SARS-CoV-2 Omicron variant'  
or is AA variant of **some** 'SARS-CoV-2 Omicron variant'))  
and ((is AA mutation of **some** 'SARS-CoV-2 Delta variant'  
or is AA variant of **some** 'SARS-CoV-2 Delta variant'))

**Query results**

Subclasses (4 of 5)

- D614G in SARS-CoV-2 S protein
- K856R in SARS-CoV-2 pp1a protein
- P314L in SARS-CoV-2 pp1b protein
- T478K in SARS-CoV-2 S protein

(A)

A: Query for AA variants shared by Delta and Omicron strains

DL query:

**Query (class expression)**

((is AA mutation of **some** 'SARS-CoV-2 Omicron variant'  
or is AA variant of **some** 'SARS-CoV-2 Omicron variant'))

**Query results**

Subclasses (45 of 46)

- A1708D in SARS-CoV-2 pp1a protein
- A63T in SARS-CoV-2 M protein
- A67V in SARS-CoV-2 S protein
- D3G in SARS-CoV-2 M protein
- D614G in SARS-CoV-2 S protein

(B)

B: Query for amino acid variants that belong to Omicron

DL query:

**Query (class expression)**

((is AA mutation of **some** 'SARS-CoV-2 Delta variant'  
or is AA variant of **some** 'SARS-CoV-2 Delta variant'))

**Query results**

Subclasses (18 of 19)

- D377Y in SARS-CoV-2 N protein
- D614G in SARS-CoV-2 S protein
- D63G in SARS-CoV-2 N protein
- D950N in SARS-CoV-2 S protein
- E156- in SARS-CoV-2 S protein

(C)

C: Query for amino acid variants that belong to Delta.

**Host-SARS-CoV-2  
gene-gene  
interaction  
network using  
SciMiner on the  
litCovid paper  
abstracts**

By: Junguk Hur,  
Oliver He

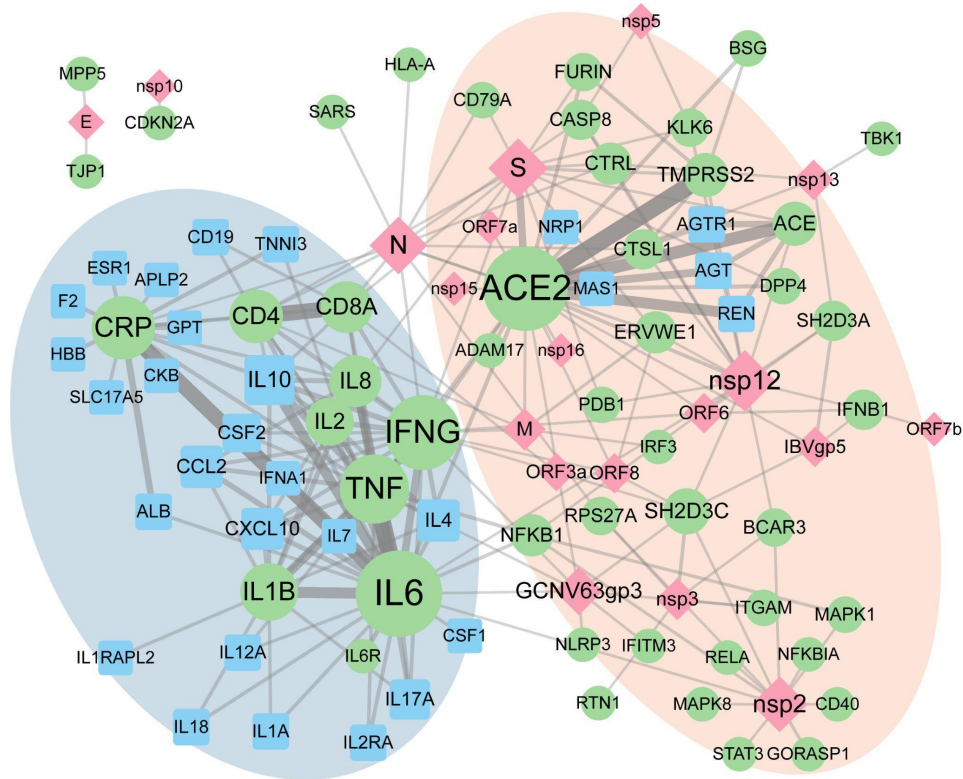


Fig. 8.

## SARS-CoV-2 drug screening based on the **drug cocktail** strategy

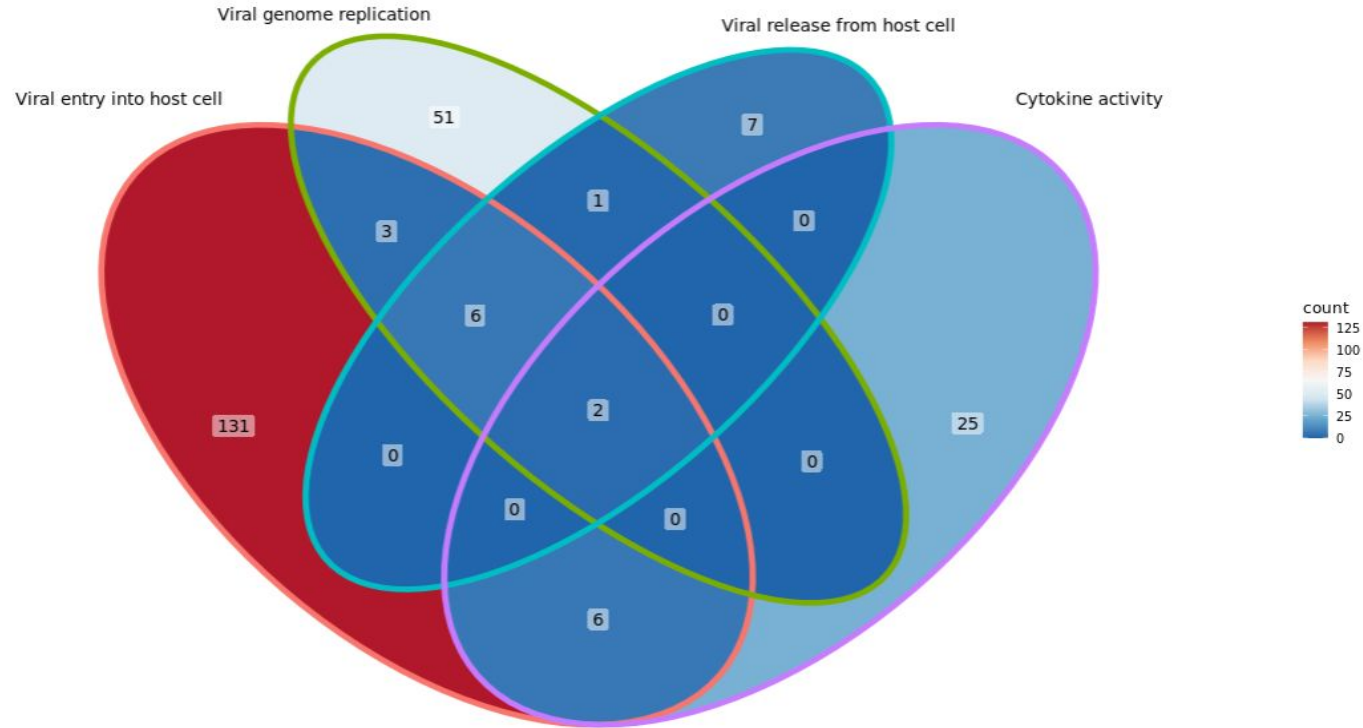


Fig. 9.



# Conclusion & Discussion

- CIDO integrates terms for coronaviruses, associated diseases, phenotypes, medical devices, treatments, vaccines, etc.
- CIDO supports computational analysis of coronavirus and disease data, evidenced by drug repurposing applications
- Representations of important coronavirus mechanisms are crucial for our progress
- **Collaborations welcome!**