Workshop on COVID-19 Ontologies (2020)

Candidate Covid-19 drugs prediction

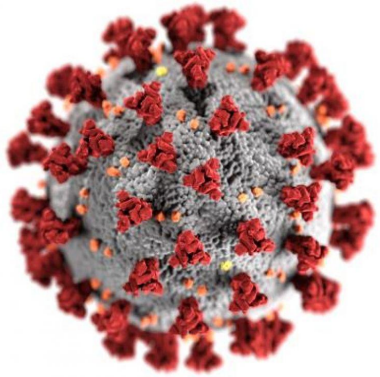
Fatima Zohra Smaili[1], Liuwei Wang[1], Yongqun He[2], Xin Gao[1], Robert Hoehndorf[1]

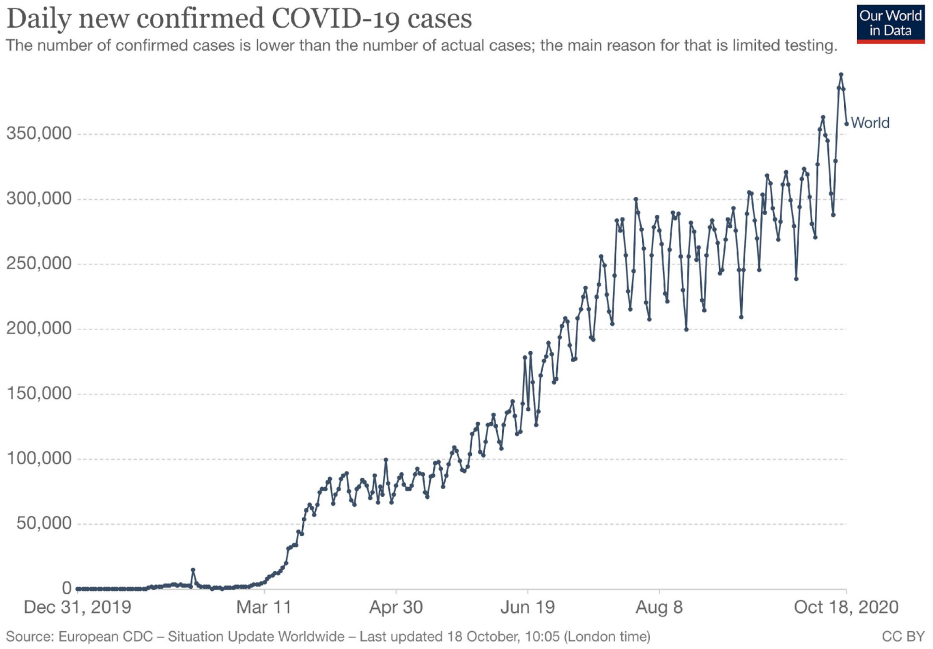
[1]King Abdullah University of Science and Technology
[2]University of Michigan Medical School

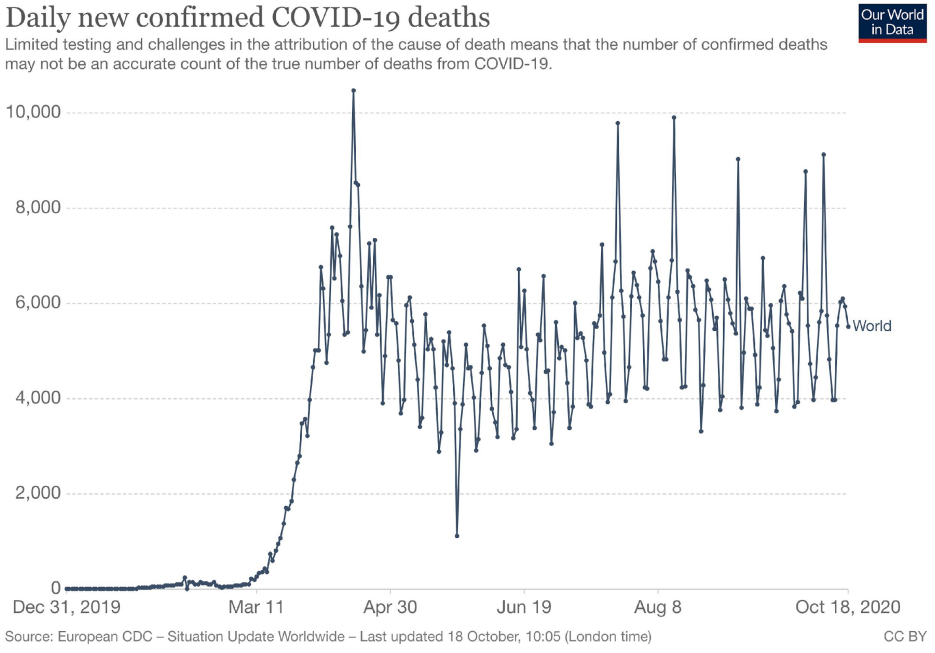
October 30, 2020

Introduction

- An efficient drug treatment for Covid-19 can dramatically reduce the severity of the disease and the death rate.
- Developing a new drug is both expensive and time consuming.
- Identify existing drugs that may be efficient in treating Covid-19.



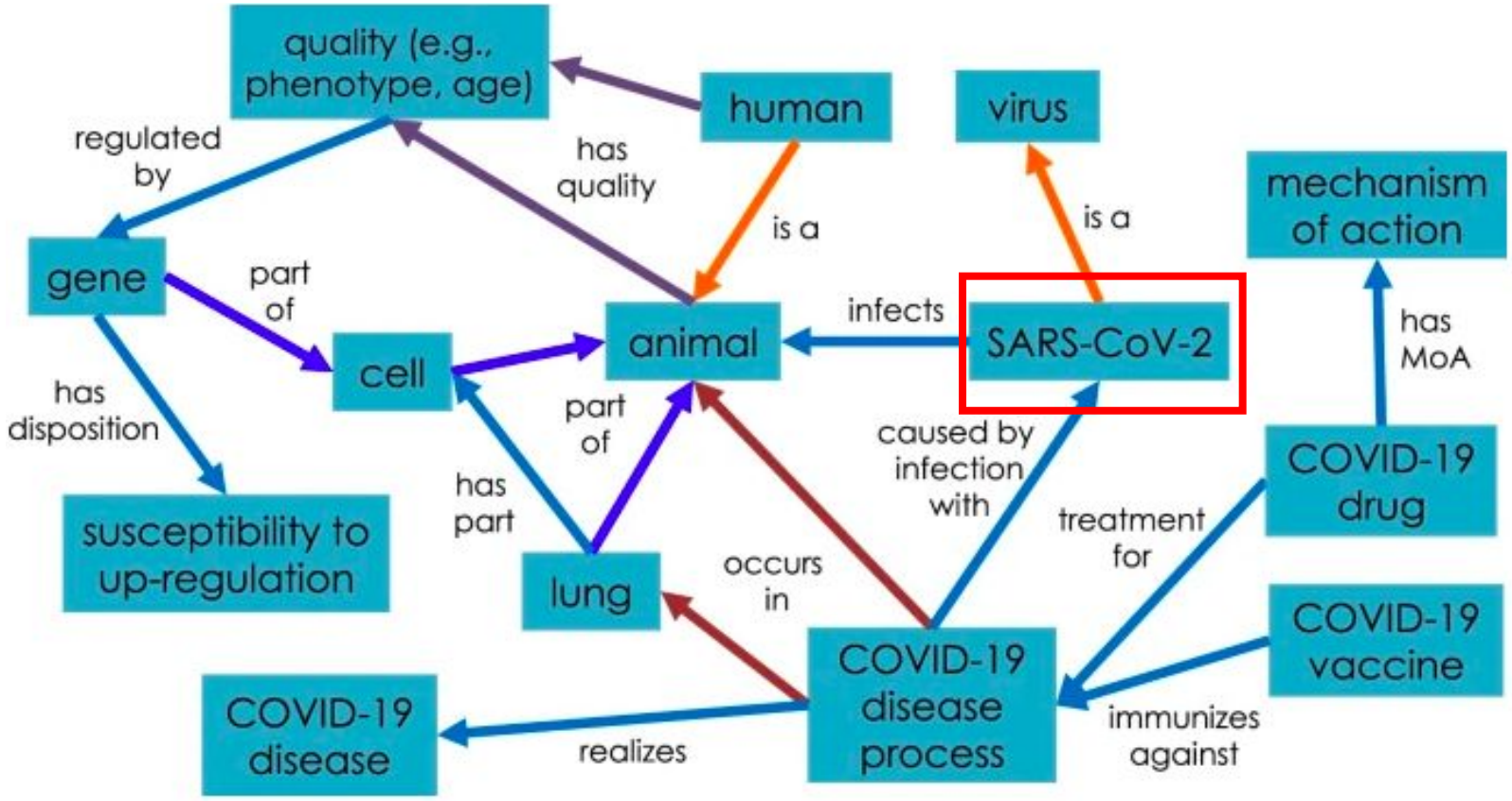




CIDO ontology

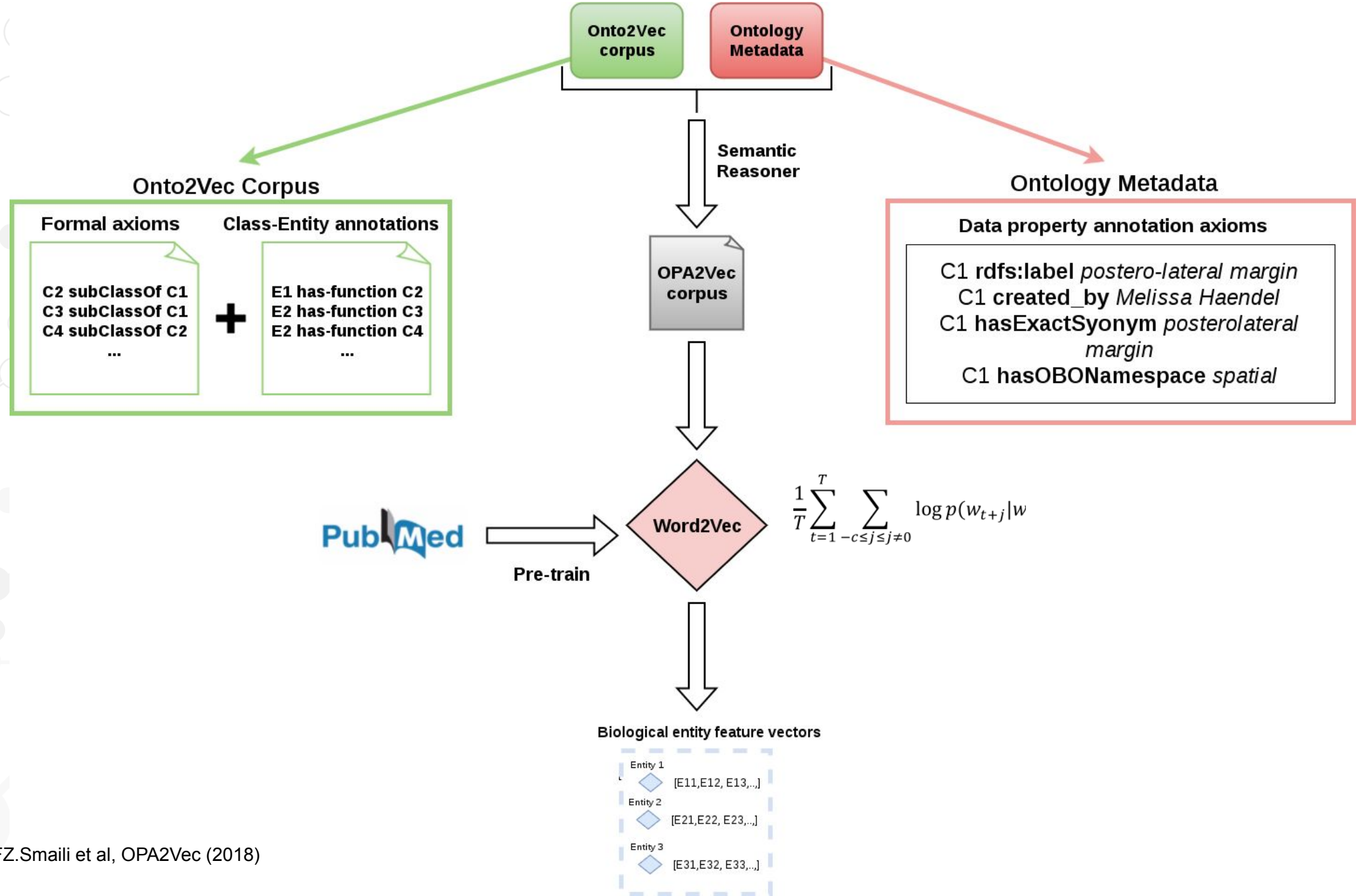
The Coronavirus Infectious Disease Ontology (CIDO): a community-based ontology that supports coronavirus disease knowledge.

*He, Y., Yu, H., Ong, E. et al. CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. Sci Data **7,** 181 (2020). https://doi.org/10.1038/s41597-020-0523-6*



OPA2Vec: Ontology-based embeddings

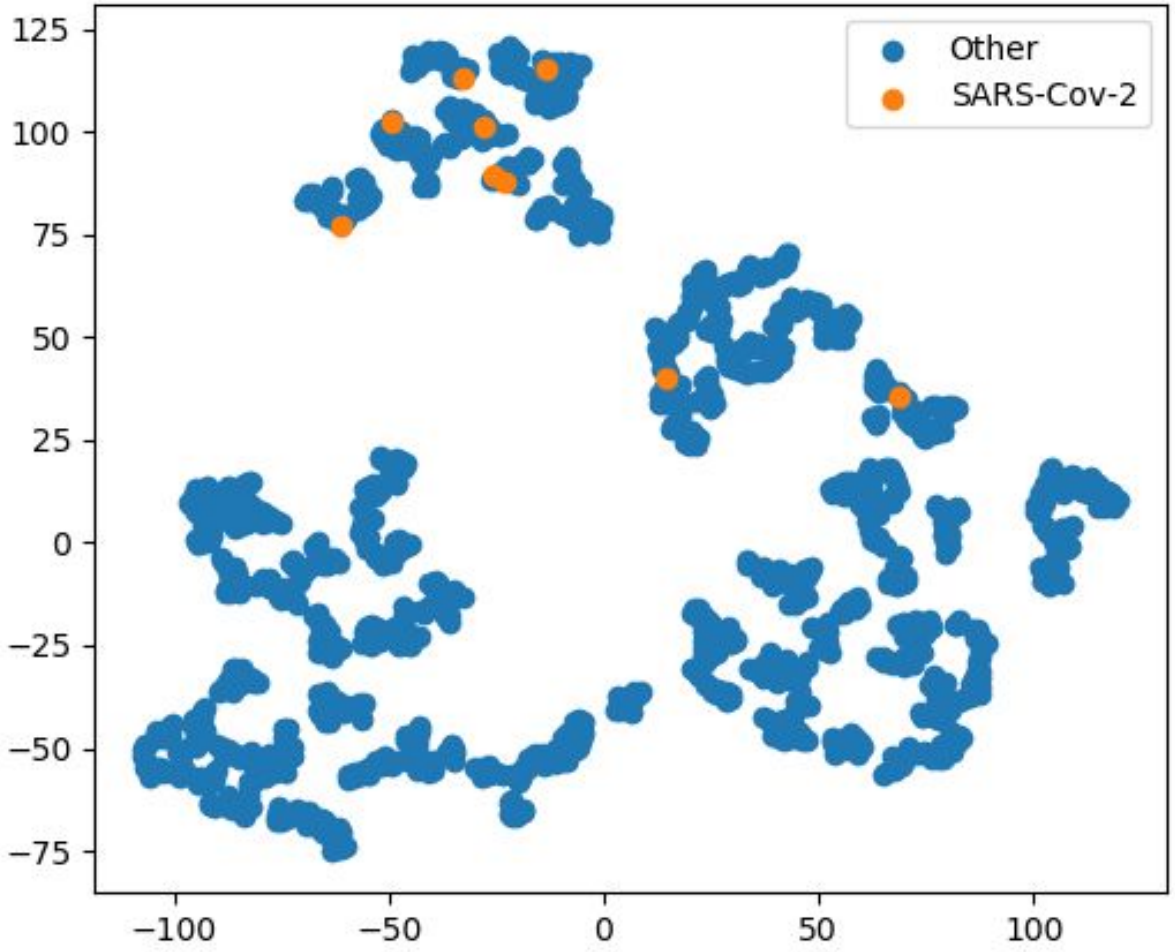
- A method that produces ontology-based representations of biological entities (proteins, diseases, drugs, etc) using:
  - Logical axioms
  - Annotation axioms
- Uses transfer learning to encode for the biomedical literature as background knowledge.



FZ.Smaili et al, OPA2Vec (2018)

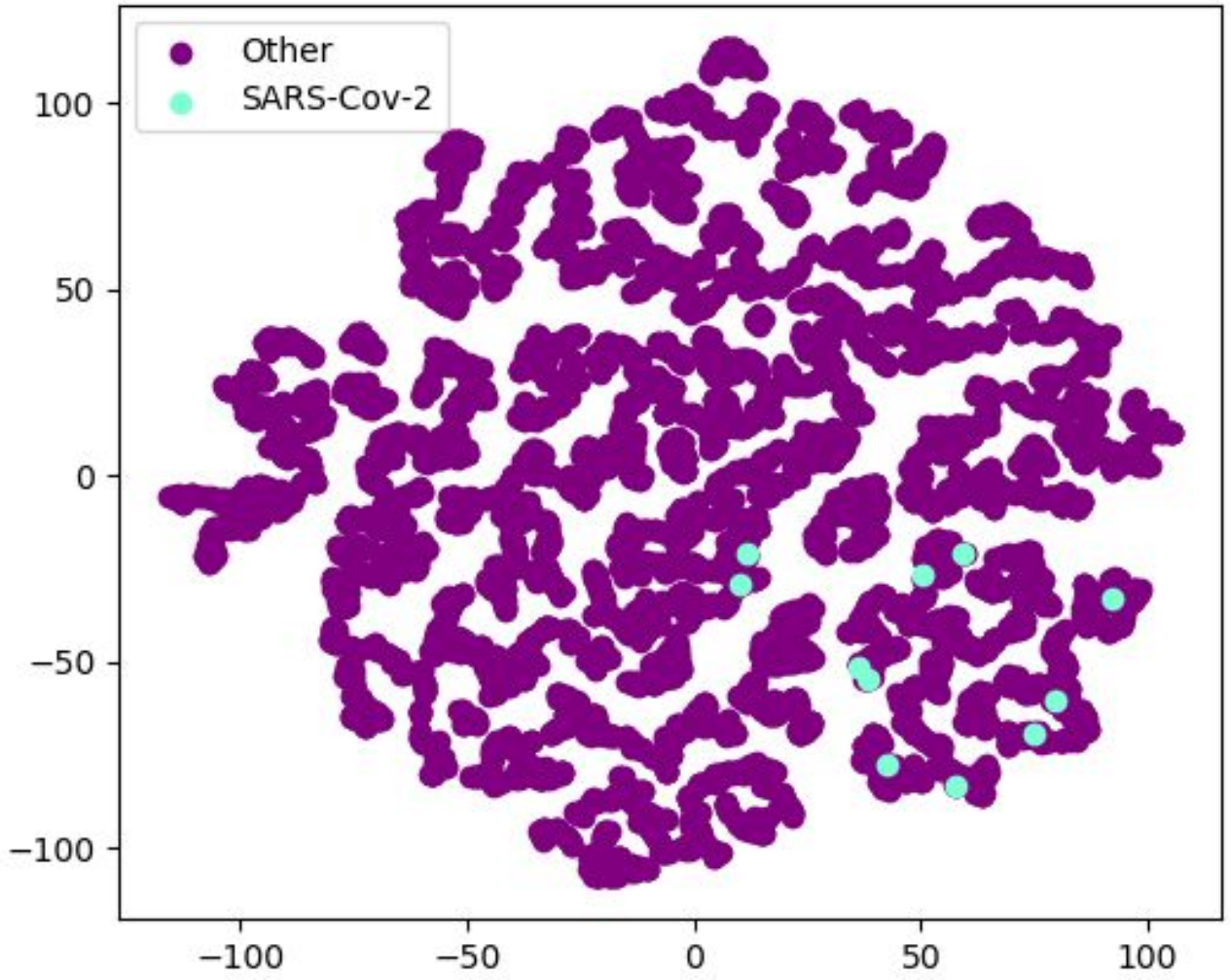
SARS-Cov-2 drugs visualization

- Visualization of drugs that inhibit the invasion of SARS-Cov-2.
- Vector representations of drugs are generated by applying OPA2Vec on CIDO ontology.



Protein targets visualization

- T-SNE based visualization of protein targets of drugs used to treat SARS-Cov-2.
- Protein vectors generated based on their Gene Ontology annotations using OPA2Vec.



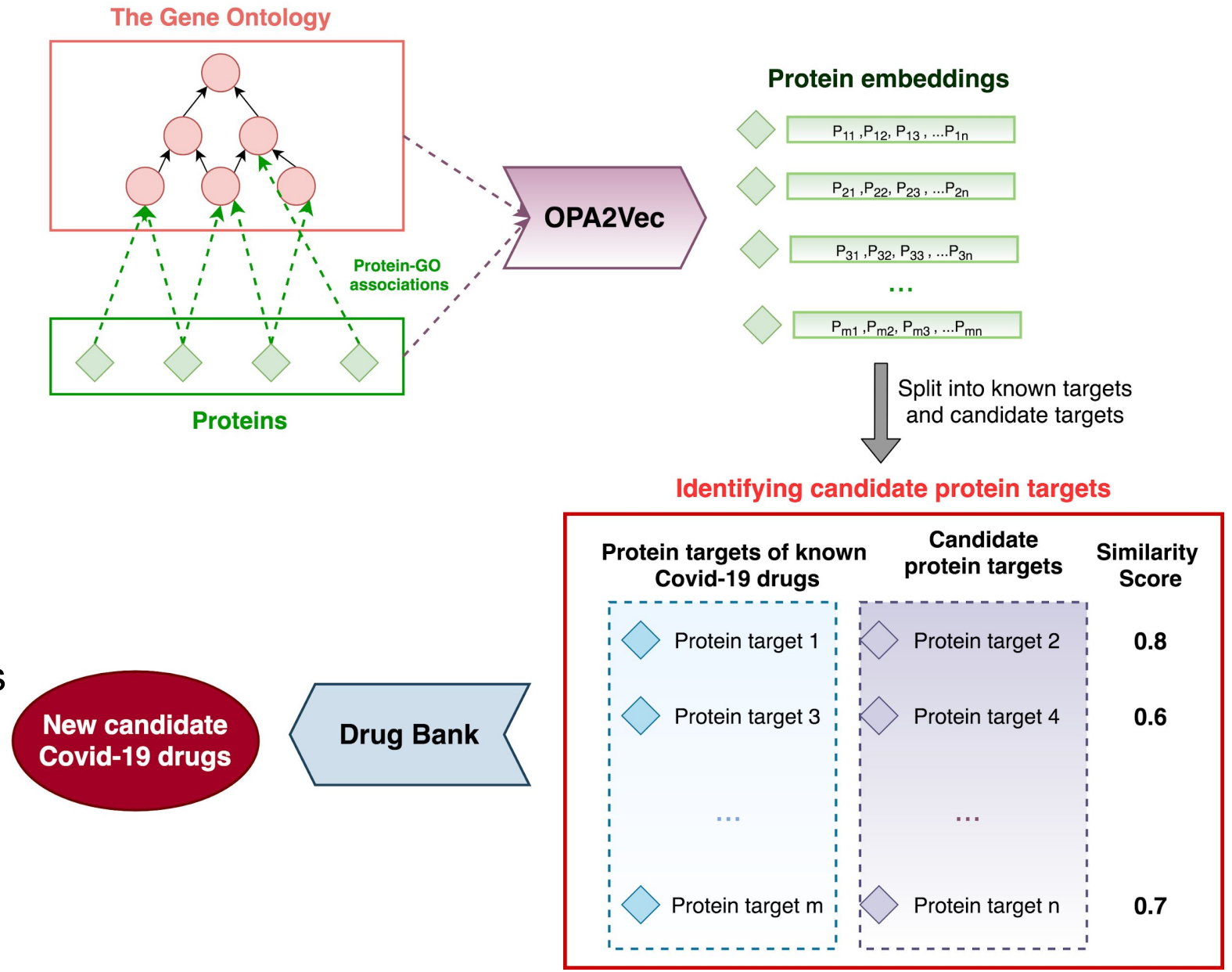
Identifying new candidate drugs

- Objective: Identify new candidate drugs for Covid-19 based on protein targets similarity.

- Method: Use OPA2Vec in combination with cosine similarity to identify similar protein targets and extract the corresponding drugs.

Workflow

- **Step 1:** Generating vector embeddings of proteins using ML-based OPA2Vec.
- **Step 2:** Calculating similarity between known protein targets of Covid-19 drugs and all other proteins to identify *candidate protein targets*.
- **Step 3**: Identifying candidate drugs based on association with the predicted protein targets using the Drug Bank.



Candidate drugs

Drug	CHEBI ID	Confidence score	Clinical trials
camostat	*CHEBI_135632*	*0.89*	*11*
gemcitabine	*CHEBI_175901*	*0.79*	*0*
tamoxifen	*CHEBI_41774*	*0.77*	*2*
imatinib	*CHEBI_45783*	*0.77*	*5*
dasatinib	*CHEBI_49375*	*0.74*	*0*
chlorpromazine	*CHEBI_3647*	*0.74*	*3*
indinavir	*CHEBI_44032*	*0.71*	*0*
azithromycin	*CHEBI_2955*	***0.69***	*114*
sirolimus	*CHEBI_9168*	*0.68*	*6*

Conclusion

- **Summary**: Identification of new candidate drugs for Covid-19 based on protein target similarity.

- **Future work**:
  - Explore the structural and sequential features of the identified drugs for further validation of their efficiency.
  - Explore drugs with different mechanisms of action their clinical use.

Thank You