Reproducible Knowledge Graph workflows for embedding chemical entities and associated biology of diseases: A use case in Mpox

Fraunhofer ITMP

Reagon Karki[1,2], Yojana Gadiya[1,2,3], Andrea Zaliani[1,2], Philip Gribbon[1,2]

[1] Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP), Schnackenburgallee 114, 22525 Hamburg, Germany

[2] Fraunhofer Cluster of Excellence for Immune-Mediated Diseases (CIMD), Theodor Stern Kai 7, 60590 Frankfurt, Germany

[3] Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, 53113 Bonn, Germany

Contact

Dr. Reagon Karki (PhD in Bioinformatics)
Researcher/Data Manager
Tel. +49 40 303764-250
reagon.karki@itmp.fraunhofer.de
Fraunhofer ITMP
Discovery Research ScreeningPort
Schnackenburgallee 114
20535, Hamburg
https://www.itmp.fraunhofer.de/en.html

Lessons learned from COVID-19

- Despite adequate expertise and technology, response was not timely[1]
- Lack of FAIR (Findable, Accessible, Interoperable, Reusable) data
- Data harmonization/integration is one of the major setbacks. Hence, reproducible workflows are required

What do we need to know about Mpox?

- Declared as a global health emergency and poses potential threat of pandemic[2]
- Mpox etiology is not known i.e., viral-host protein interaction, virus replication biology, etc.
- Identify first line of drugs (short-term)
- Develop/design/repurpose drugs for Mpox (long-term)

The role of Knowledge Graphs (KGs)

- KGs can efficiently embed various scientific data such as chemical/biological entities, assays, pathways, etc.
- KGs enable FAIR data
- KGs unravel potential disease mechanism(s)[3]
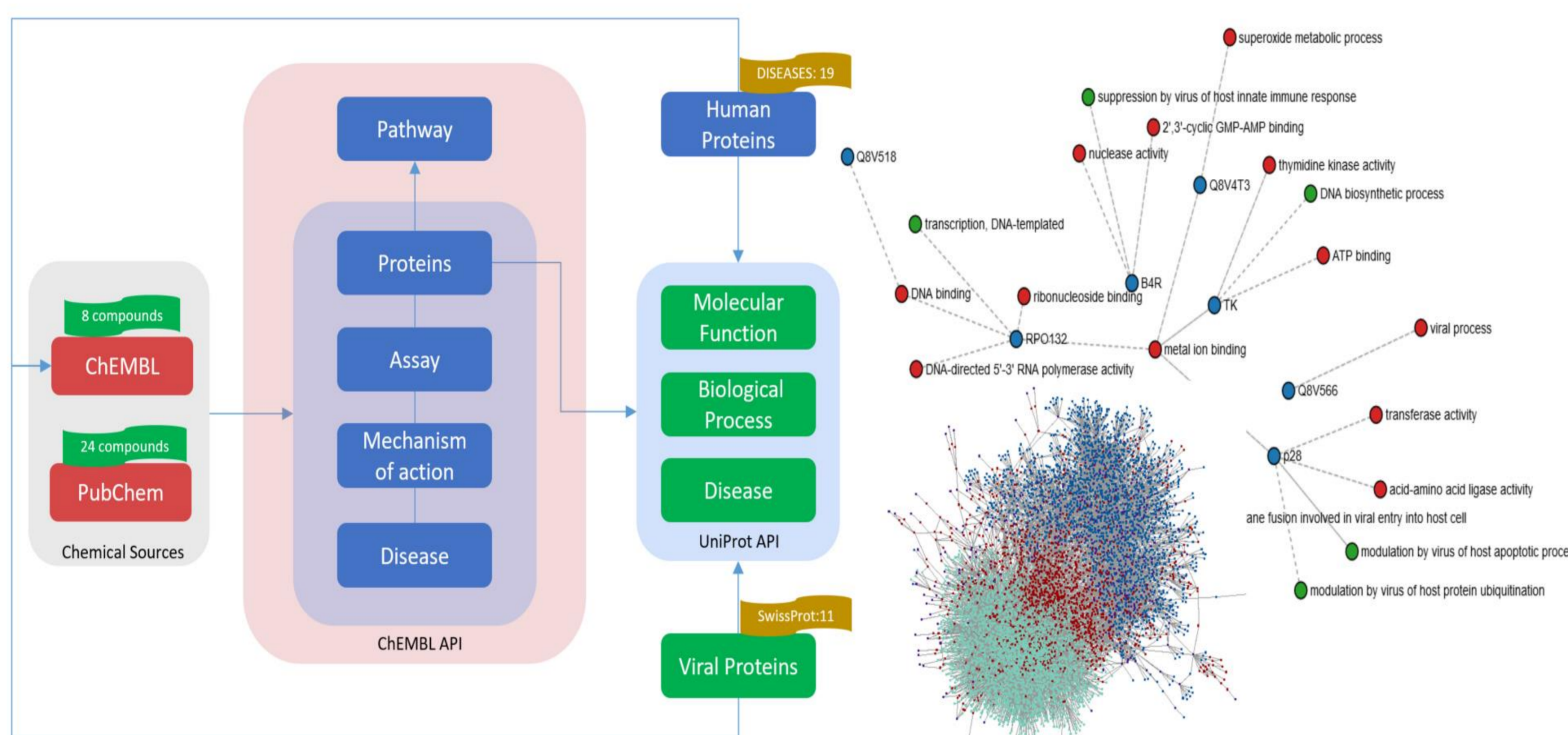- KGs serve as the basis for downstream analyses

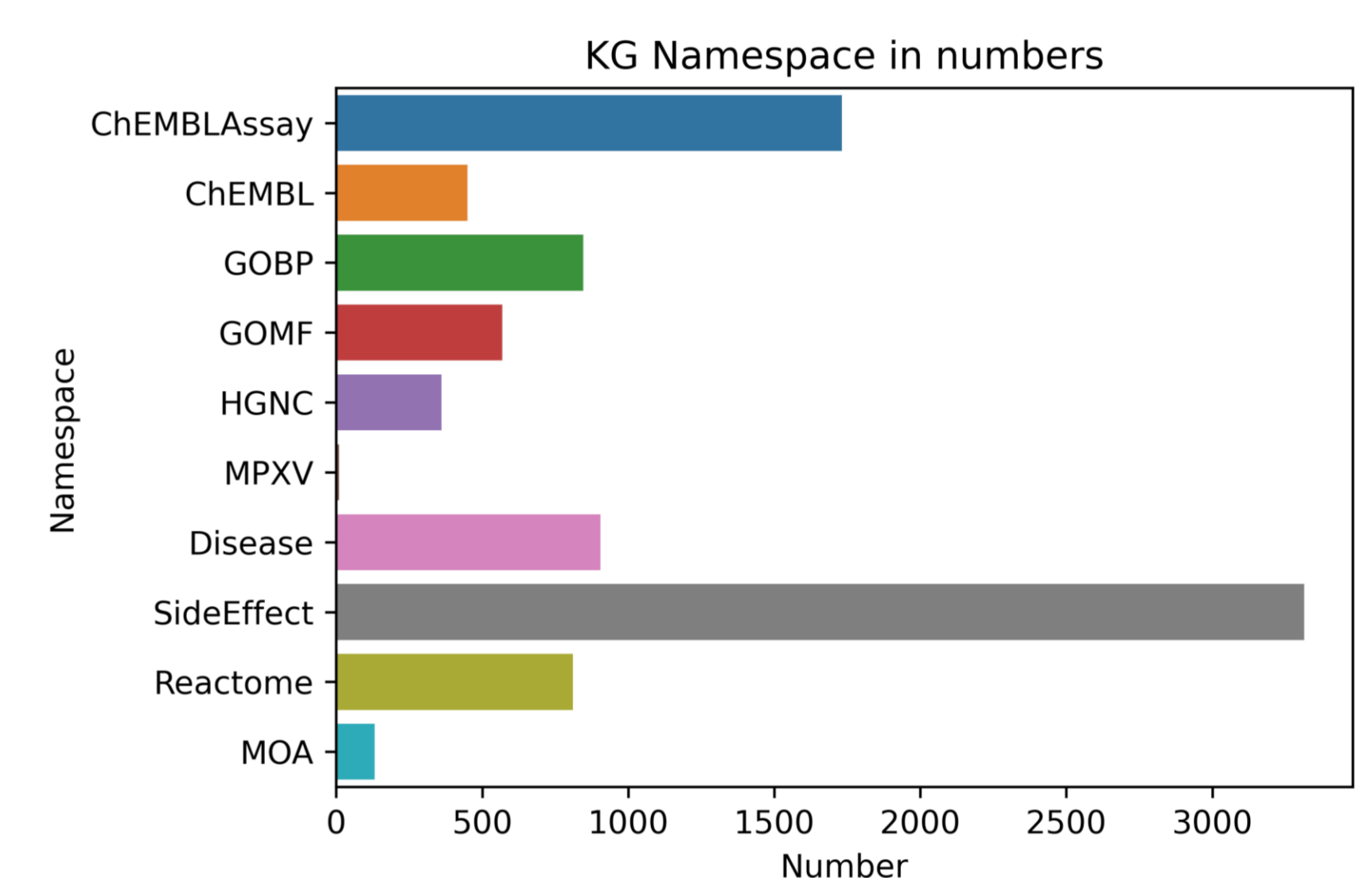*Figure 1: A schematic representation of the KG workflow (left) and visualization of the KG (right)*

Figure 2: A summary of various entities and their numbers in KG

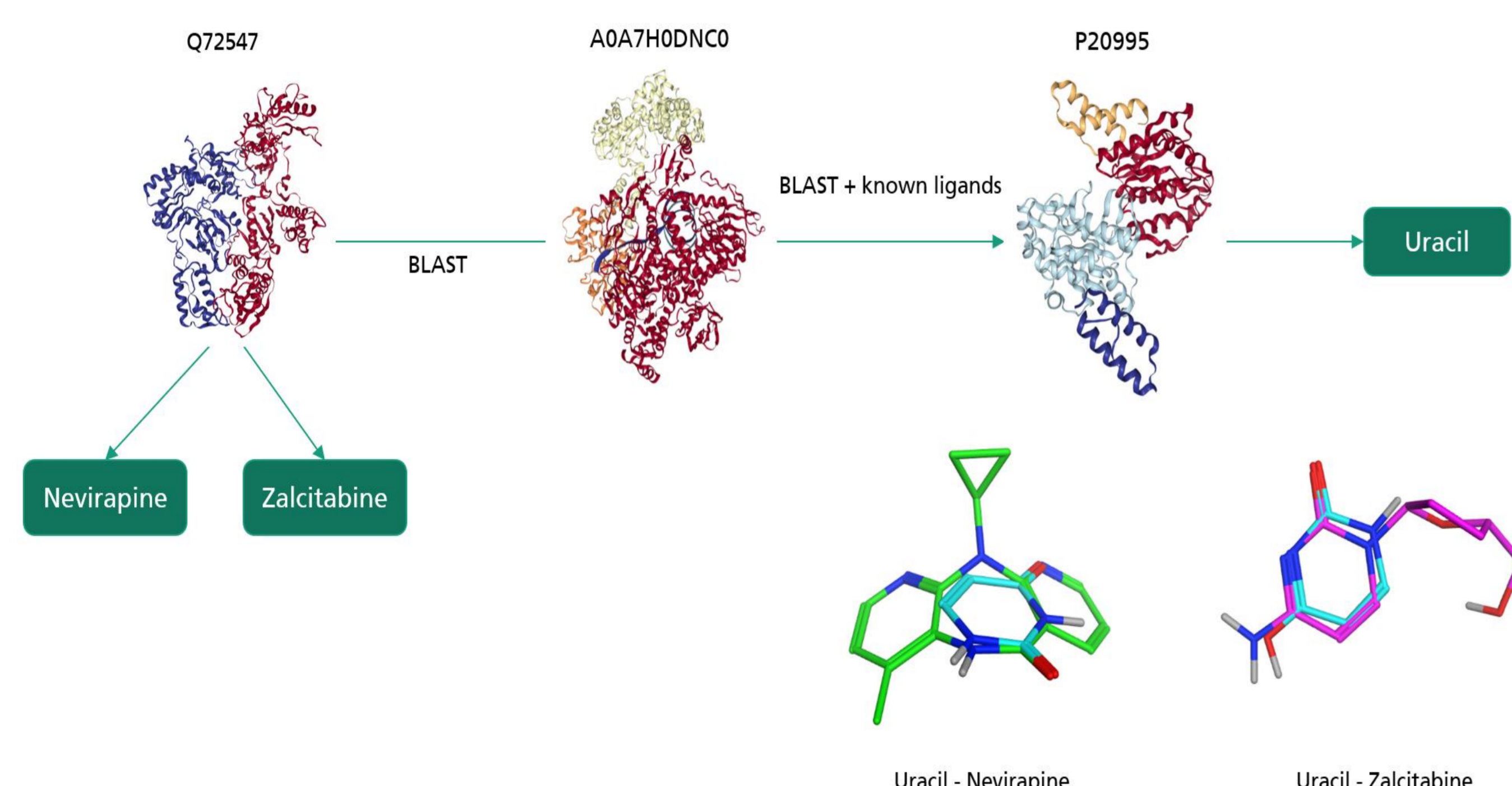## KG Analysis: Identification of drug repurposing candidates

Figure 3: Downstream chemoinformatics analyses of chemicals identified in KG

Target Groups

- Researchers looking for biomedical information related to macromolecular structures
- Researchers interested in active compounds against target proteins
- Researchers performing assay-based experiments
- Researchers in pre-clinical drug discovery projects
- Chemoinformaticians

Way Forward

- A reproducible KG workflow in place for generating phenotype-chemotype of disease of interests
- Collaborate with experts from other domains and align resources
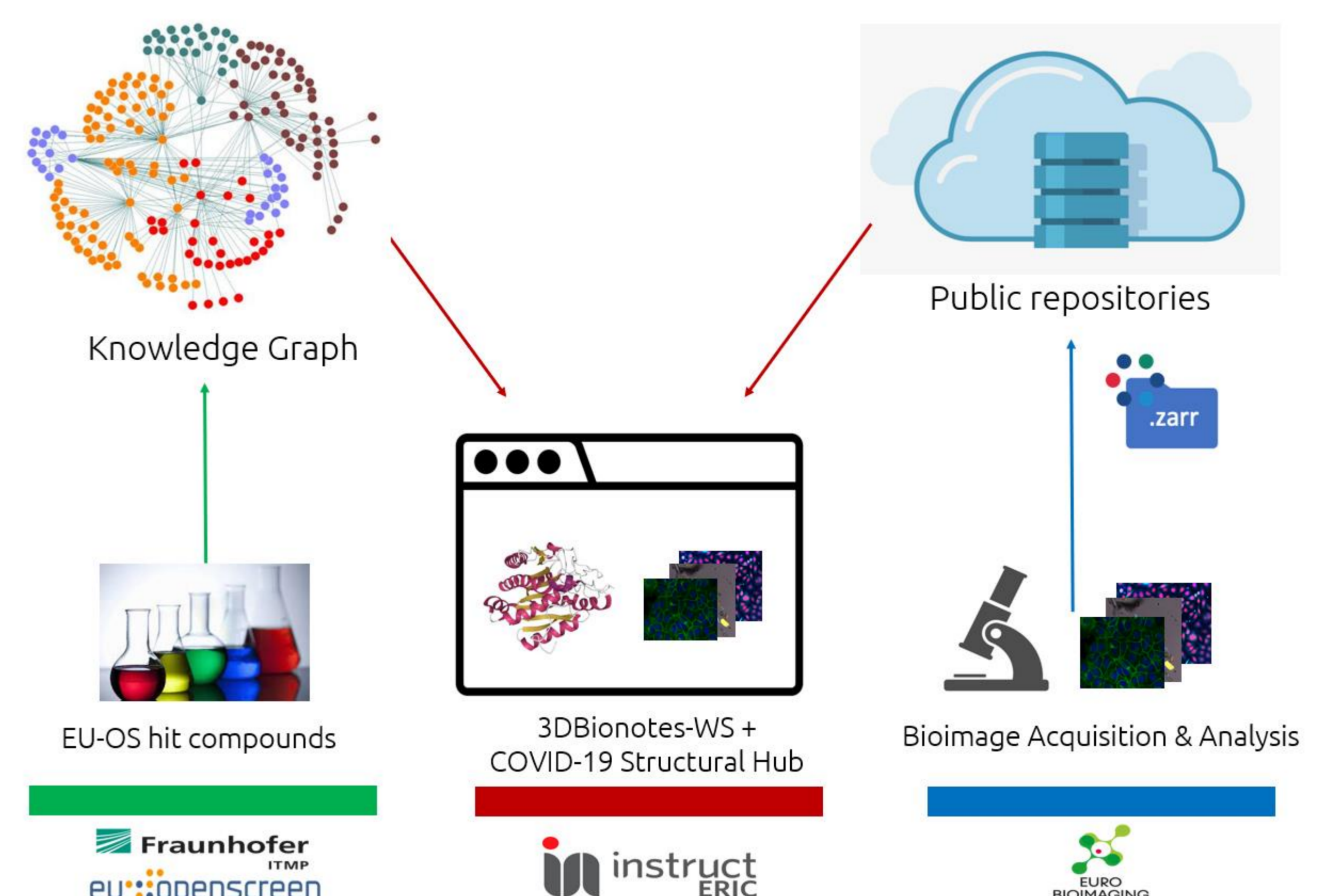- Develop/improve downstream analysis pipelines

Figure 4: Integrating data from different domains in COVID-19 and beyond

Resource/Data Availability

- Mpox KG: https://github.com/Fraunhofer-ITMP/mpox-kg
- Fragment Screening: https://github.com/Fraunhofer-ITMP/COVID-NMR-KG
- COVID19 KG: https://github.com/Fraunhofer-ITMP/BY-COVID-KG
- Anti-Microbial Resistance KG: https://github.com/Fraunhofer-ITMP/AMR-KG

Related Publications:

- Mpox Knowledge Graph: A comprehensive representation embedding chemical entities and associated biology of Mpox (https://doi.org/10.1093/bioadv/vbad045)
- Comprehensive Fragment Screening of the SARS-CoV-2 Proteome Explores Novel Chemical Space for Drug Development (https://doi.org/10.1002/anie.202205858)

Acknowledgments

References

1 Villa, S., Lombardi, A., Mangioni, D., Bozzi, G., Bandera, A., Gori, A., & Raviglione, M. C. (2020). The COVID-19 pandemic preparedness or lack thereof: from China to Italy. *Global Health & Medicine*.
2 https://www.ecdc.europa.eu/en/Mpox-outbreak
3 Domingo-Fernández, Daniel, et al. "COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology." *Bioinformatics* 37.9 (2021): 1332-1334.

KG Demo

Funded by:

BY-COVID   EOSC Future