Biological systems are described and modeled through combining various data types1. The availability of ‘omics datasets, such as in Ensembl, NCBI’s GEO repository, as well as molecular profiling of patients in the clinic, provide opportunities to elucidate regulatory networks2 and drug mechanisms3 promoting personalized medicine4,5. Additionally, integrating patient clinical characteristics into these systems can improve our ability to treat disease, understand physiology, and enhance clinical trial methodologies6,7 at both the individual and population level.

There is limited data infrastructure for integrating patient clinical characteristics with associated molecular samples. The Health Information Technology for Economic and Clinical Health Act (HITECH Act), signed into law in 2009, was passed to motivate the implementation of electronic health records (EHR) and supporting technology in the United States8. However, clinical diagnostics assays and molecular profiling has only recently become efficient, cheap, and interpretable for clinical practice. The OHDSI Common Data Model (CDM) provides a standardized infrastructure that allows population and patient-level analyses to promote evidence-based medical practice9. Clinicians and researchers are able to make specific queries and run comprehensive, reproducible analyses over many sites following this CDM10. Researchers have begun to develop extensions such as for integrating patient genomics data in the CDM5 and further characterizing cancer subtypes11. Unfortunately, these infrastructures are specific for a disease domain and do not integrate patient data with data from molecular profiling. While tools to visualize and explore patient data have been developed12,13, there are no tools for utilizing the OMOP CDM to integrate molecular profiling of both patients in the clinic and individuals from population-based studies.

We present ROMOPOmics, an R package that extends the OMOP CDM for characterizing associated patient molecular data. We show how consortium datasets from TCGA and experimental datasets such as T-cell epigenomic data are integrated into the OMOP CDM extension by ROMOPOmics. We then discuss the opportunities and research directions that become available using this data infrastructure. ROMOPOmics seeks to lay the groundwork for standardizing patient clinical and molecular profiling data using the established OMOP framework while remaining adaptable enough to incorporate new and changing sequencing techniques, ensuring that the clinical and research communities can remain as up-to-date with the changing data landscape as possible.

1. Ideker, T., Galitski, T. & Hood, L. A NEW APPROACH TO DECODING LIFE: Systems Biology. *Annu. Rev. Public Health* **27**, 297–322 (2006).

2. Knaack, S. A., Siahpirani, A. F. & Roy, S. A Pan-Cancer Modular Regulatory Network Analysis to Identify Common and Cancer-Specific Network Components. *Cancer Inform.* **13s5**, CIN.S14058 (2014).

3. Woo, J. H. *et al.* Elucidating Compound Mechanism of Action by Network Perturbation Analysis. *Cell* **162**, 441–451 (2015).

4. Biankin, A. V. The road to precision oncology. *Nat. Genet.* **49**, 320–321 (2017).

5. Gómez-López, G., Dopazo, J., Cigudosa, J. C., Valencia, A. & Al-Shahrour, F. Precision medicine needs pioneering clinical bioinformaticians. *Brief. Bioinform.* **20**, 752–766 (2019).

6. Bielekova, B., Vodovotz, Y., An, G. & Hallenbeck, J. How Implementation of Systems Biology into Clinical Trials Accelerates Understanding of Diseases. *Front. Neurol.* **5**, 1–9 (2014).

7. Assmus, H. E., Herwig, R., Cho, K.-H. & Wolkenhauer, O. Dynamics of biological systems: role of systems biology in medical research. *Expert Rev. Mol. Diagn.* **6**, 891–902 (2006).

8. Blumenthal, D. & Tavenner, M. The “Meaningful Use” Regulation for Electronic Health Records. *N. Engl. J. Med.* **363**, 501–504 (2010).

9. Park, R. W. Sharing clinical big data while protecting confidentiality and security: Observational health data sciences and informatics. *Healthc. Inform. Res.* **23**, 1–3 (2017).

10. Kashyap, M. *et al.* Development and validation of phenotype classifiers across multiple sites in the observational health data sciences and informatics network. *J. Am. Med. Informatics Assoc.* **27**, 877–883 (2020).

11. Belenkaya, R. *et al.* Standardized Observational Cancer Research Using the OMOP CDM Oncology Module. *Stud. Health Technol. Inform.* **264**, 1831–1832 (2019).

12. Glicksberg, B. S. *et al.* ROMOP: a light-weight R package for interfacing with OMOP-formatted electronic health record data. *JAMIA Open* **2**, 10–14 (2019).

13. Glicksberg, B. S. *et al.* PatientExploreR: An extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model. *Bioinformatics* **35**, 4515–4518 (2019).