The Journal of Open Source Software

Drug Informatics Studio: An Ontology-based Integrative Research Environment for Clinical Informatics

Shyh-Huei Chen[1], **Ryan Barnard**[1], **Lynn S. Huang**[1], **Barry I. Freedman**[2], **and Jing Su**[1]

1 Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA **2** Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

DOI: 10.21105/joss.02218

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Submitted: 20 May 2020
Published: 22 May 2020

Summary

The Drug Informatics Studio is an R package designed to provide easy-to-use functions for semantic data mining of the drug prescription history in electronic medical records (EMR) according to the common biomedical ontologies.

A patient's "pharmacy history" provides a unique and essential information domain that reflects health status, delivered healthcare, treatment responses, comorbidities, and other important clinical features of the patient. Extracting drug information from EMR and other real-world data and integrating this information in the context of knowledge bases of drugs and diseases, known as *drug informatics*, plays a pivotal role in medical informatics and forms the foundation of clinical decision support systems. However, a huge gap remains between the rich and fast-growing EMR data of drug history and the meaningful and reliable representation of prescribed drugs of an individual patient. The major challenge lies between the complexity and dynamics of the real-world drug history data and the underdeveloped informatics approaches to comprehensively and meaningfully represent the health information behind drug prescriptions and to use drug-related health information in clinical research and patient care.

Drug ontologies comprehensively annotate drugs from multiple perspectives, from mechanisms to associated diseases to chemical ingredients to known drug interactions. Commonly used drug ontologies and controlled vocabularies include RxNorm, National Drug Code (NDC), the National Drug File – Reference Terminology (NDF-RT, now the Medication Reference Terminology: MED-RT) (Brown et al., 2004), First DataBank (FDB) MedKnowledge and OrderKnowledge, the Anatomical Therapeutic Chemical (ATC) Classification System, and others. Each drug ontology systematically organizes such annotations with the relations among them in the form of a heterogeneous directed acyclic graph. Drug ontologies, as well as controlled vocabularies, are then cross-mapped from one ontology to the other through semantic relations. Together, drug ontologies and controlled vocabularies as well as the cross-mapping between them form a *pan-ontology drug semantic network*. Such a network provides the foundation for multi-faceted, extensible, and scalable representations of health information represented by drugs and the intrinsic similarities among drugs. Thus, mapping drug records onto this drug semantic network and using semantic similarity metrics for EMR data mining are able to address the challenge of meaningfully using real-world drug records in clinical research and patient care.

To address this challenge in drug informatics and bridge the gap between the wealth of EMR data and the urgent needs in clinical informatics research, we developed an ontology-based, integrated research environment, the *Drug Informatics Studio* (Chen et al., n.d.) Figure 1, to: 1) systematically annotate drugs, encounters, and patients; 2) define similarity metrics among

drugs, encounters, and patients; 3) subtype patients according to their drug history, and 4) support modeling for precision medicine and clinical decision support. The Drug Informatics Studio is composed of a pan-ontology network to provide comprehensive ontology support, ontology-indexed drug EMRs to represent drug records using this network, and an informatics toolbox to enable clinical research using ontology-based drug information. Currently, the Drug Informatics Studio covers the National Drug File – Reference Terminology (NDF-RT) for medications as well as associated Unified Medical Language System (UMLS) Metathesaurus ontologies and terminologies, such as RxNorm, that support the NDF-RT ontologies. Annotations are mainly based on NDF-RT's drug annotation concepts, including mechanisms of action (MoA), physiologic effects (PE), and pathophysiologic as well as certain non-disease physiologic states (Disease). The enrichment of annotation concepts enables clinical research to define similarity metrics among drugs, subtype patients according their drug history, and support the modeling for precision medicine and clinical decision support. A set of functions are provided for searching similar drugs using ontological concepts, identifying patients that use similar drugs, and visualizing which annotation concepts are enriched among similar patients.

We used simulated data derived from the EMRs of prevalent Wake Forest patients with diabetes to demonstrate the functionalities of the Drug Informatics Studio. This demonstration cohort of 1,000 virtual patients were simulated from a WFBH cohort includes 122,632 patients, 342,347 encounters, and 15,918,328 drug records. We performed annotations as well as drug-based clustering of clinical encounters. We further profiled the enriched clinical features in these newly discovered clusters. We demonstrated how to use the Drug Informatics Studio to systematically annotate drugs, index patients' drug information in the EMR, define similarity metrics among drugs, subtype patients according to their drug history, and support the modeling for precision medicine and clinical decision support. Such simulated EMR data are provided to demonstrate the functionalities (for example, MoA concept enrichment analysis shown in Figure 2 and Figure 3) of the Drug Informatics Studio.

Currently, the U.S. Study to Protect Brain Health Through Lifestyle Intervention to Reduce Risk (U.S. POINTER)(Baker et al., 2019, p. @RN4) is using the Drug Informatics Studio to support the EMR-based participant screening for targeted recruitment.

Figures

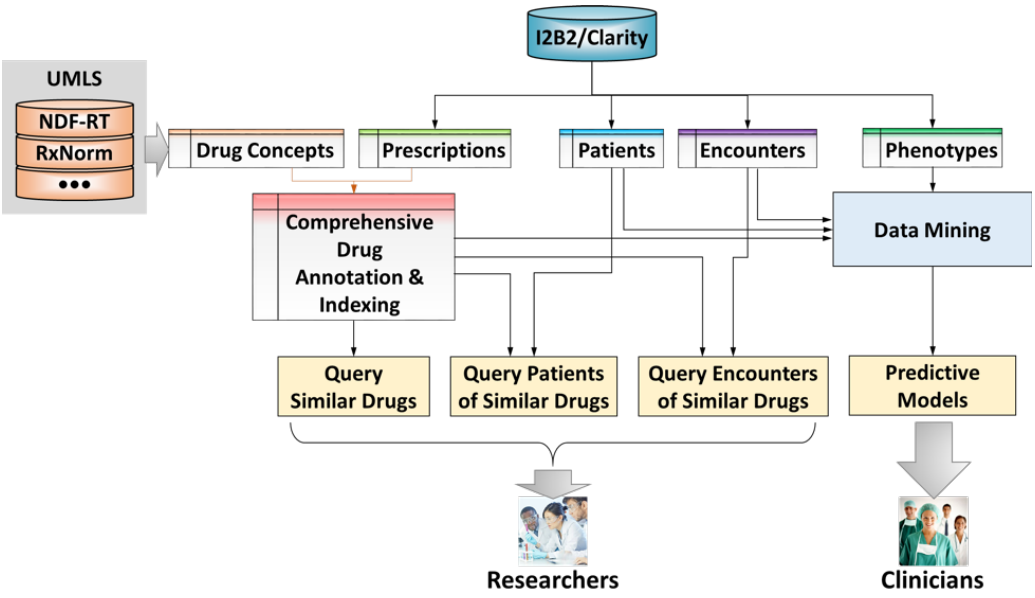

Figure 1: Schema of the Drug Informatics Studio

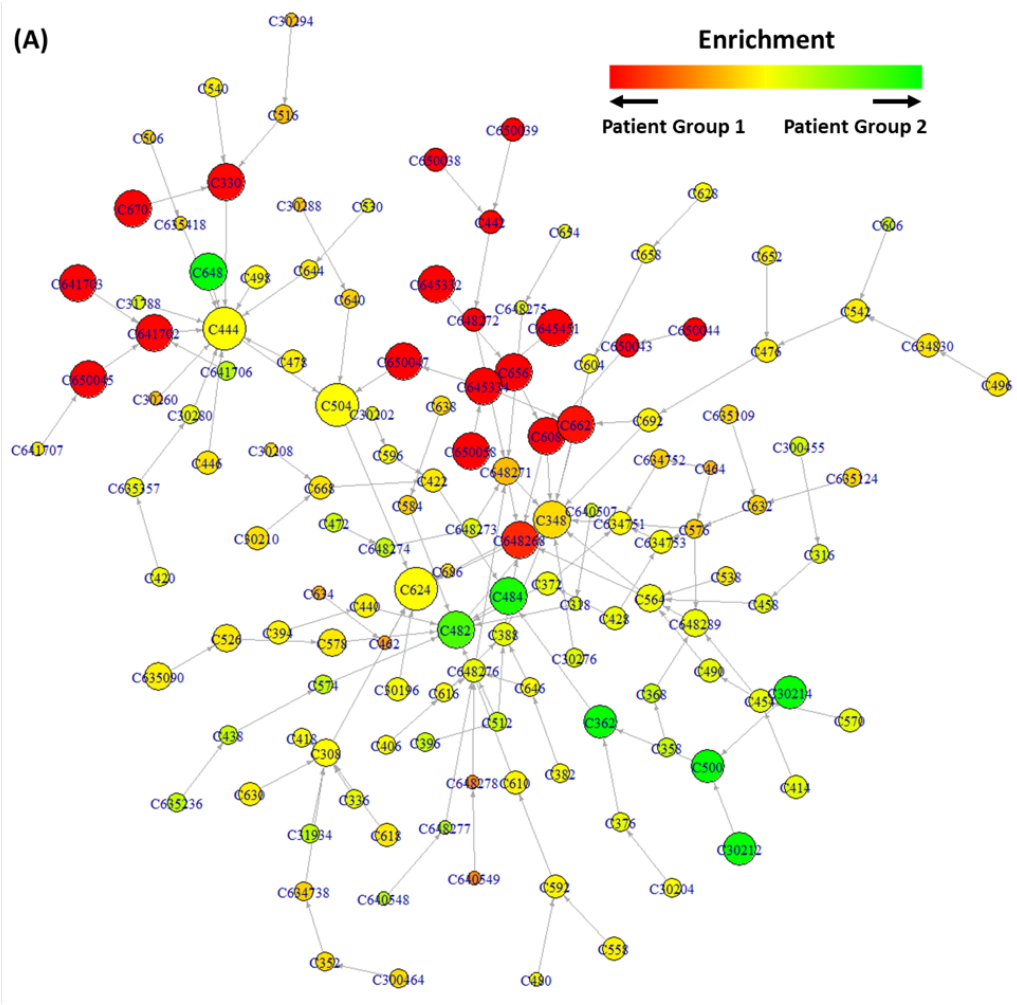

Figure 2: See Figure 3 Caption

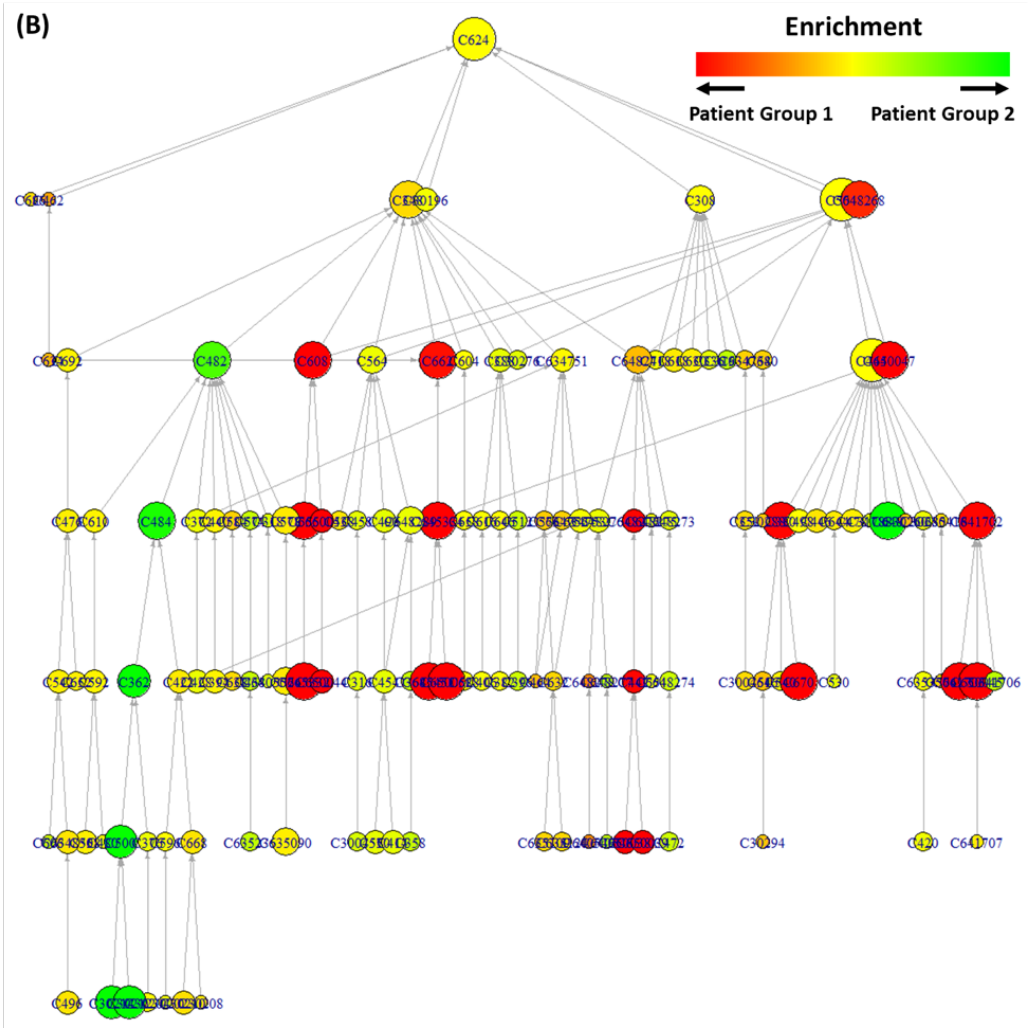

Figure 3: Demonstration of MoA concepts differentially enriched in Patient Group 1 and Patient Group 2 in tree layout (A) or large-graph layout (B). The graph is a subgraph of the MoA ontology, with each node as an MoA concept and each edge pointing from a more specific MoA concept toward to a more general one. The color represents the relative enrichment of an MoA concept in Patient Group 1 (red) vs. Patient Group 2 (green). The size of a node represents the logarithmic frequency of the MoA concept in the overall dataset. MoA concepts of low frequency (less than 20) are not shown.

Acknowledgement

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001420. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Baker, L. D., Beavers, D. P., Cleveland, M., Day, C. E., Decarli, C., Espeland, M. A., Tomaszewski-Farias, S. E., et al. (2019). US pointer: STUDY design and launch.

Alzheimer's & Dementia, *15*, P1262–P1263. Journal Article.

Brown, S. H., Elkin, P. L., Rosenbloom, S. T., Husser, C., Bauer, B. A., Lincoln, M. J., Carter, J., et al. (2004). VA national drug file reference terminology: A cross-institutional content coverage study. *Stud Health Technol Inform*, *107*(Pt 1), 477–81. Journal Article. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15360858

Chen, S.-H., Huang, L. S., Barnard, R., Hsu, F.-C., Freedman, B. I., & Su, J. (n.d.). Drug informatics studio: An ontology-based integrative research environment for clinical informatics. In *AMIA*. Conference Proceedings.

Su, J., Chen, S.-H., Huang, L. S., Hsu, F.-C., Baker, L. D., & Espeland, M. A. (2018). EMR-based pragmatic participant recruitment for the u.s. Pointer clinical trial. *Alzheimer's & Dementia*, *14*, P567–P567. Journal Article. doi:10.1016/j.jalz.2018.06.610