

# Charles Tapley Hoyt, Ph.D. - Publications

Author lists have been omitted to reduce visual clutter. The following are noted:

- † signifies first or co-first authorship
- ‡ signifies last or senior authorship

1. **Eliater: A Python package for estimating outcomes of perturbations in biomolecular networks.** *Bioinformatics*, 2024. doi:10.1093/bioinformatics/btae527
2. †**The O3 guidelines: open data, open code, and open infrastructure for sustainable curated scientific resources.** *Scientific Data*, 2024. doi:10.1038/s41597-024-03406-w
3. **An open source knowledge graph ecosystem for the life sciences.** *Scientific Data*, 2024. doi:10.1038/s41597-024-03171-w
4. ‡**ptwt - The PyTorch Wavelet Toolbox.** *JMLR*, 2024
5. (in preparation) **SeMRA: Assembly and Reasoning over Semantic Mappings at Scale for Biomedical Data Integration.** *in preparation*, 2024
6. (in preparation) **Improving reproducibility of cheminformatics workflows with ‘chembl-downloader’.** *in preparation for J. Chem. Inf.*, 2024
7. (in preparation) **Causal inference with  $Y_0$ .** *in preparation for JOSS*, 2024
8. **A simple standard for ontological mappings 2023: updates on data model, collaborations and tooling.** *OM 2023*
9. **The Human Phenotype Ontology in 2024: phenotypes around the world.** *NAR*, 2023. doi:10.1093/nar/gkad1005
10. †**Improving reproducibility and reusability in the Journal of Cheminformatics.** *J Cheminform*, 2023. doi:10.1186/s13321-023-00730-y
11. **Democratizing knowledge representation with BioCypher.** *Nature Biotechnology*, 2023. doi:10.1038/s41587-023-01848-y
12. **Optimal adjustment sets for causal query estimation in partially observed biomolecular networks.** *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad270
13. †**Prediction and curation of missing biomedical identifier mappings with Biomappings.** *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad130
14. (preprint) **Experimental design for causal query estimation in partially observed biomolecular networks.** *arXiv*, 2022. arxiv:2210.13423
15. †**Unifying the identification of biomedical entities with the Bioregistry.** *Scientific Data*, 2022. doi:10.1038/s41597-022-01807-3
16. **A Simple Standard for Ontological Mappings 2022: Updates of data model and outlook.** *OM 2022*
17. **A review of biomedical datasets relating to drug discovery: a knowledge graph perspective.** *Brief Bioinform*, 2022. doi:10.1093/bib/bbac404
18. **ChemicalX: A Deep Learning Library for Drug Pair Scoring.** *KDD*, 2022. doi:10.1145/3534678.3539023
19. **Integrating multi-omics data reveals function and therapeutic potential of deubiquitinating enzymes.** *eLife*, 2022. doi:10.7554/elife.72879
20. **Understanding the Performance of Knowledge Graph Embeddings in Drug Discovery.** *Artificial Intelligence in the Life Sciences*, 2022. doi:10.1016/j.ailsci.2022.100036
21. **A Simple Standard for Sharing Ontological Mappings (SSSOM).** *Database*, 2022. doi:10.1093/database/baac035
22. **Gilda: biomedical entity text normalization with machine-learned disambiguation as a service.** *Bioinformatics Advances*, 2022. doi:10.1093/bioadv/vbac034
23. **ProtSTonKGs: A Sophisticated Transformer Trained on Protein Sequences, Text, and Knowledge Graphs.** *SWAT4HCLS 2022*. ceur-ws:3127:13
24. (preprint) **Mondo: Unifying diseases for the world, by the world.** *medRxiv*, 2022. doi:10.1101/2022.04.13.22273750

25. **Do-calculus enables estimation of causal effects in partially observed biomolecular pathways.** *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac251
26. (preprint) **†A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs.** *arXiv*, 2022. arxiv:2203.07544
27. **PyBioPAX: biological pathway exchange in Python.** *JOSS*, 2022. doi:10.21105/joss.04136
28. (preprint) **An Open Challenge for Inductive Link Prediction on Knowledge Graphs.** *arXiv*, 2022. arxiv:2203.01520
29. **STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs.** *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac001
30. **Ontology Development Kit: a toolkit for building, maintaining, and standardising biomedical ontologies.** *Database*, 2022. doi:10.1093/database/baac087
31. **Bringing Light Into the Dark: A Large-scale Evaluation of Knowledge Graph Embedding Models Under a Unified Framework.** *TPAMI*, 2021. doi:10.1109/tpami.2021.3124805
32. **The role of metadata in reproducible computational research.** *Patterns*, 2021. doi:10.1016/j.patter.2021.100322
33. **‡CLEP: a hybrid data- and knowledge-driven framework for generating patient representations.** *Bioinformatics*, 2021. doi:10.1093/bioinformatics/btab340
34. **A Systems Biology Approach for Hypothesizing the Effect of Genetic Variants on Neuroimaging Features in Alzheimer's Disease.** *JAD*, 2021. doi:10.3233/jad-201397
35. **†PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings.** *JMLR*, 2021. arxiv:2007.14175
36. **Leveraging Structured Biological Knowledge for Counterfactual Inference: A Case Study of Viral Pathogenesis.** *IEEE TBData*, 2021. doi:10.1109/tbdata.2021.3050680
37. (preprint) **†Extension of Roles in the ChEBI Ontology.** *ChemRxiv*, 2020. doi:10.26434/chemrxiv.12591221
38. **The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST).** *Bioinformatics*, 2020. doi:10.1093/bioinformatics/btaa622
39. **GuiltyTargets: Prioritization of Novel Therapeutic Targets with Deep Network Representation Learning.** *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020. doi:10.1109/tcbb.2020.3003830
40. **PS4DR: a multimodal workflow for identification and prioritization of drugs based on pathway signatures.** *BMC Bioinf.*, 2020. doi:10.1186/s12859-020-03568-5
41. **Identifying the parametric occurrence of multiple steady states for some biological networks.** *Journal of Symbolic Computation*, 2020. doi:10.1016/j.jsc.2019.07.008
42. **A Computational Approach for Mapping Heme Biology in the Context of Hemolytic Disorders.** *Frontiers in Bioengineering and Biotechnology*, 2020. doi:10.3389/fbioe.2020.00074
43. **The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling.** *Frontiers in Genetics*, 2019. doi:10.3389/fgene.2019.01203
44. **Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm.** *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-3101-1
45. **The KEEN Universe.** *ISWC 2019*. doi:10.1007/978-3-030-30796-7\_1
46. **Predicting Missing Links Using PyKEEN.** *ISWC 2019*. ceur-ws:2456:64
47. **RatVec: A General Approach for Low-dimensional Distributed Vector Representations via Domain-specific Rational Kernels.** *LWDA 2019*
48. **BioKEEN: a library for learning and evaluating biological knowledge graph embeddings.** *Bioinformatics*, 2019. doi:10.1093/bioinformatics/btz117
49. **PathMe: merging and exploring mechanistic pathway knowledge.** *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-2863-9
50. (preprint) **†Integration of Structured Biological Data Sources using Biological Expression Language.** *bioRxiv*, 2019. doi:10.1101/631812 bioRxiv:631812

51. **ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases.** *NPJ Syst Biol Appl.*, 2019. doi:10.1038/s41540-018-0078-8
52. **†Re-curation and rational enrichment of knowledge graphs in Biological Expression Language.** *Database*, 2019. doi:10.1093/database/baz068
53. **Challenges of Integrative Disease Modeling in Alzheimer's Disease.** *Frontiers in molecular biosciences*, 2019. doi:10.3389/fmolb.2019.00158
54. **†BEL Commons: an environment for exploration and analysis of networks encoded in Biological Expression Language.** *Database*, 2018. doi:10.1093/database/bay126
55. **BEL2ABM: agent-based simulation of static models in Biological Expression Language.** *Bioinformatics*, 2018. doi:10.1093/bioinformatics/bty107
56. **†PyBEL: a Computational Framework for Biological Expression Language.** *Bioinformatics*, 2018. doi:10.1093/bioinformatics/btx660
57. **†A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities.** *Database*, 2018. doi:10.1093/database/bay050
58. **A Case Study on the Parametric Occurrence of Multiple Steady States.** *ISSAC 2017*. doi:10.1145/3087604.3087622
59. **Repurposing human PDE4 inhibitors for neglected tropical diseases: design, synthesis and evaluation of cilomilast analogues as Trypanosoma brucei PDEB1 inhibitors.** *Bioorg Med Chem Lett*, 2014. doi:10.1016/j.bmcl.2014.07.063

## Invited Presentations (Selected)

1. **Modern software development practice with Python.** *May Institute* (May 7, 2024)
2. **Assembly and inference over semantic mappings to support the NFDI Terminology Service.** *TS4NFDI Community Workshop* (April 9, 2024)
3. **Assembly of Domain Knowledge at Scale in Biomedicine and Beyond.** *Harvard Medical School - Laboratory of Systems Pharmacology Meeting* (January 19, 2024)
4. **Machine-assisted integration of data and knowledge at scale to support biomedical discovery.** *NIH BISTI Seminar* (December 7, 2023)
5. **Democratizing Biocuration, or, How I Learned to Love the Drive-by Curation.** *International Society of Biocuration Annual General Meeting* (October 18, 2023)
6. **Improving ontology interoperability with Biomappings.** *OBO Academy - Monarch Training Series* (September 19, 2023)
7. **Modern prefix management with the Bioregistry and 'curies'.** *OBO Academy - Monarch Training Series* (September 5, 2023)
8. **Axiomatizing Chemical Roles.** *Ontologies4Chem Workshop 2022* (September 7, 2022)
9. **Modern Scientific Software Development Practice in Python.** *May Institute* (May 20, 2022)
10. **Knowledge Graph Embedding with PyKEEN in 2022.** *Knowledge Graph Conference (KGC 2022)* (May 5, 2022)
11. **The Biopragmatics Stack: Biomedical and Chemical Semantics for Humans.** *Machine-Actionable Data Interoperability for Chemical Sciences (MADICES)* (February 8, 2022)
12. **Reusable Science in Python.** *May Institute* (July 28, 2024)
13. **Current Issues in Theory, Reproducibility, and Utility of Graph Machine Learning in the Life Sciences.** *Graph Machine Learning in Industry* (September 23, 2021)
14. **Perspectives on Knowledge Graph Embedding Models in/out of Biomedicine.** *AstraZeneca* (April 6, 2021)
15. **Introduction to the Biological Expression Language and the Rational Enrichment Workflow.** *CoronaWhy* (May 6, 2020)
16. **Applications of Knowledge Graphs in Drug Discovery.** *Computational Drug Discovery Group, University of Leiden* (November 5, 2019)
17. **Generation and Application of Biomedical Knowledge Graphs.** *Harvard Medical School* (July 19, 2019)

## Talks and Posters (Selected)

1. **Assembly and Reasoning over Semantic Mappings at Scale.** *Biocuration 2024* (March 7, 2024)
2. **Introduction to WPCI 2023.** *Winter 2023 Workshop on Prefixes, CURIEs, and IRIs* (November 27, 2023)
3. **Standardization of chemical prefixes, CURIEs, URIs, and semantic mappings.** *Ontologies4Chem Workshop 2023* (October 12, 2023)
4. **Improving the reproducibility of cheminformatics workflows with chembl-downloader.** *RDKit User Group Meeting 2023* (September 21, 2023)
5. **Promoting the longevity of curated scientific resources through open code, open data, and public infrastructure.** *Biocuration 2023* (April 26, 2023)
6. **Using dashboards to monitor ontology standardisation and community activity.** *Ontology Summit 2023* (February 15, 2023)
7. **Introduction to WPCI 2022.** *2022 Workshop on Prefixes, CURIEs, and IRIs* (December 5, 2022)
8. **The Bioregistry, CURIEs, and OBO Community Health.** *International Conference on Biomedical Ontology (ICBO)* (September 26, 2022)
9. **Closing the Semantic Gap: Identifying Missing Mappings and Merging Equivalent Concepts to Support Knowledge Graph Assembly.** *Harvard Medical School - Sorger Lab Meeting* (August 1, 2022)
10. **A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs.** *Graph Learning Benchmarks (GLB 2022)* (April 26, 2022)
11. **Introduction to WPCI 2021.** *2021 Workshop on Prefixes, CURIEs, and IRIs* (October 29, 2021)
12. **Biomappings: Community Curation of Mappings between Biomedical Entities.** *4th Session of the International Society of Biocuration 2021 Virtual Conference* (October 5, 2021; poster)
13. **The Bioregistry: A Metaregistry for Biomedical Entities.** *12th International Conference on Biomedical Ontologies* (September 17, 2021)
14. **Future Directions for WikiPathway Meta-curation.** *WikiPathways Developers Conference Call* (January 6, 2021)
15. **The Biological Expression Language and PyBEL in 2020.** *COVID-19 Disease Map Community Meeting* (July 10, 2020)
16. **Maintenance and Enrichment of Disease Maps in Biological Expression Language.** *4th Disease Maps Community Meeting* (October 4, 2019; poster)
17. **Identifying Drug Repositioning Candidates using Representation Learning on Heterogeneous Networks.** *The Eighth Joint Sheffield Conference on Chemoinformatics* (June 19, 2019; poster)
18. **The PyBEL Ecosystem in 2018.** *OpenBEL Community Meeting* (May 14, 2018)
19. **From Knowledge Assembly to Hypothesis Generation.** *Bio-IT World* (April 22, 2018)
20. **Knowledge Assembly in Systems and Networks Biology.** *Bio-IT World* (April 23, 2018; poster)
21. **The Human Brain Pharmacome: An Overview.** *3rd European Conference on Translational Bioinformatics* (April 17, 2018; poster)
22. **Gene Set Analysis using Phenotypic Screening Data.** *Research, Innovation and Scholarship Expo 2015* (April 9, 2015; poster)