Charles Tapley Hoyt, Ph.D. - Publications

Author lists have been omitted to reduce visual clutter. There are a total of 62 manuscripts (53 peer-reviewed, 8 pre-printed/submitted, and 1 in preparation). The following are noted:

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

1. †**Assembly and reasoning over semantic mappings at scale for biomedical data integration**. *Bioinformatics*, 2025. doi:10.1093/bioinformatics/btaf542
2. †**Improving reproducibility of cheminformatics workflows with chembl-downloader**. *JOSS*, 2025. doi:10.21105/joss.08844
3. (preprint) †**Causal identification with Y0**. *arXiv*, 2025. doi:10.48550/arxiv.2508.03167
4. (preprint) **The Cell Ontology in the age of single-cell omics**. *arXiv*, 2025. doi:10.48550/arxiv.2506.10037
5. †**Computational tools and data integration to accelerate vaccine development: challenges, opportunities, and future directions**. *Front Immunol*, 2025. doi:10.3389/fimmu.2025.1502484
6. (in preparation) †**Assembly and application of coherent biomedical lexica**. *in preparation for Database*, 2025
7. (submitted) **OmniPath: integrated knowledgebase for multi-omics analysis**. *submitted to Nucleic Acids Research*, 2025. biorxiv:2025.09.11.675512
8. (submitted) **VO: The Vaccine Ontology**. *submitted to Scientific Data*, 2025. biorxiv:2025.08.12.669998
9. (submitted) ‡**More Rigorous Software Engineering Would Improve Reproducibility in Machine Learning Research**. *submitted to TMLR*, 2025. arxiv:2502.00902
10. (in preparation) **Teaching RDM in a smart advanced inorganic lab course and its provision in the DALIA platform**. *in preparation for Nature Computational Science*, 2025
11. **Eliater: A Python package for estimating outcomes of perturbations in biomolecular networks**. *Bioinformatics*, 2024. doi:10.1093/bioinformatics/btae527
12. †**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
13. **An open source knowledge graph ecosystem for the life sciences**. *Scientific Data*, 2024. doi:10.1038/s41597-024-03171-w
14. ‡**ptwt - The PyTorch Wavelet Toolbox**. *JMLR*, 2024
15. **A simple standard for ontological mappings 2023: updates on data model, collaborations and tooling**. *OM 2023*
16. **The Human Phenotype Ontology in 2024: phenotypes around the world**. *NAR*, 2023. doi:10.1093/nar/gkad1005
17. †**Improving reproducibility and reusability in the Journal of Cheminformatics**. *J Cheminform*, 2023. doi:10.1186/s13321-023-00730-y
18. **Democratizing knowledge representation with BioCypher**. *Nature Biotechnology*, 2023. doi:10.1038/s41587-023-01848-y
19. **Optimal adjustment sets for causal query estimation in partially observed biomolecular networks**. *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad270
20. †**Prediction and curation of missing biomedical identifier mappings with Biomappings**. *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad130
21. (preprint) **Experimental design for causal query estimation in partially observed biomolecular networks**. *arXiv*, 2022. arxiv:2210.13423
22. †**Unifying the identification of biomedical entities with the Bioregistry**. *Scientific Data*, 2022. doi:10.1038/s41597-022-01807-3
23. **A Simple Standard for Ontological Mappings 2022: Updates of data model and outlook**. *OM 2022*
24. **A Review of Biomedical Datasets Relating to Drug Discovery: A Knowledge Graph Perspective**. *Brief Bioinform*, 2022. doi:10.1093/bib/bbac404

25. **ChemicalX: A Deep Learning Library for Drug Pair Scoring**. *KDD*, 2022. doi:10.1145/3534678.3539023
26. **Integrating multi-omics data reveals function and therapeutic potential of deubiquitinating enzymes**. *eLife*, 2022. doi:10.7554/elife.72879
27. **Understanding the Performance of Knowledge Graph Embeddings in Drug Discovery**. *Artificial Intelligence in the Life Sciences*, 2022. doi:10.1016/j.ailsci.2022.100036
28. **A Simple Standard for Sharing Ontological Mappings (SSSOM)**. *Database*, 2022. doi:10.1093/database/baac035
29. **Gilda: biomedical entity text normalization with machine-learned disambiguation as a service**. *Bioinformatics Advances*, 2022. doi:10.1093/bioadv/vbac034
30. **ProtSTonKGs: A Sophisticated Transformer Trained on Protein Sequences, Text, and Knowledge Graphs**. *SWAT4HCLS 2022*. ceur-ws:3127:13
31. (preprint) **Mondo: Unifying diseases for the world, by the world**. *medRxiv*, 2022. doi:10.1101/2022.04.13.22273750
32. **Do-calculus enables estimation of causal effects in partially observed biomolecular pathways**. *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac251
33. (preprint) †**A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs**. *arXiv*, 2022. arxiv:2203.07544
34. **PyBioPAX: biological pathway exchange in Python**. *JOSS*, 2022. doi:10.21105/joss.04136
35. (preprint) **An Open Challenge for Inductive Link Prediction on Knowledge Graphs**. *arXiv*, 2022. arxiv:2203.01520
36. **STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs**. *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac001
37. **Ontology Development Kit: a toolkit for building, maintaining, and standardising biomedical ontologies**. *Database*, 2022. doi:10.1093/database/baac087
38. **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
39. **The role of metadata in reproducible computational research**. *Patterns*, 2021. doi:10.1016/j.patter.2021.100322
40. ‡**CLEP: a hybrid data- and knowledge-driven framework for generating patient representations**. *Bioinformatics*, 2021. doi:10.1093/bioinformatics/btab340
41. **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
42. †**PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings**. *JMLR*, 2021. arxiv:2007.14175
43. **Leveraging Structured Biological Knowledge for Counterfactual Inference: A Case Study of Viral Pathogenesis**. *IEEE TBDATA*, 2021. doi:10.1109/tbdata.2021.3050680
44. (preprint) †**Extension of Roles in the ChEBI Ontology**. *ChemRxiv*, 2020. doi:10.26434/chemrxiv.12591221
45. **The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST)**. *Bioinformatics*, 2020. doi:10.1093/bioinformatics/btaa622
46. **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
47. **PS4DR: a multimodal workflow for identification and prioritization of drugs based on pathway signatures**. *BMC Bioinf.*, 2020. doi:10.1186/s12859-020-03568-5
48. **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
49. **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
50. **The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling**. *Frontiers in Genetics*, 2019. doi:10.3389/fgene.2019.01203
51. **Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation**

amplitude (CMPA) algorithm. *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-3101-1

52. **The KEEN Universe - An Ecosystem for Knowledge Graph Embeddings with a Focus on Reproducibility and Transferability**. *ISWC 2019*. doi:10.1007/978-3-030-30796-7$_1$
53. **Predicting Missing Links Using PyKEEN**. *ISWC 2019*. ceur-ws:2456:64
54. **RatVec: A General Approach for Low-dimensional Distributed Vector Representations via Domain-specific Rational Kernels**. *LWDA 2019*
55. **BioKEEN: a library for learning and evaluating biological knowledge graph embeddings**. *Bioinformatics*, 2019. doi:10.1093/bioinformatics/btz117
56. **PathMe: merging and exploring mechanistic pathway knowledge**. *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-2863-9
57. (preprint) †**Integration of Structured Biological Data Sources using Biological Expression Language**. *bioRxiv*, 2019. doi:10.1101/631812 biorxiv:631812
58. **ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases**. *NPJ Syst Biol Appl.*, 2019. doi:10.1038/s41540-018-0078-8
59. †**Re-curation and rational enrichment of knowledge graphs in Biological Expression Language**. *Database*, 2019. doi:10.1093/database/baz068
60. **Challenges of Integrative Disease Modeling in Alzheimer's Disease**. *Frontiers in molecular biosciences*, 2019. doi:10.3389/fmolb.2019.00158
61. †**BEL Commons: an environment for exploration and analysis of networks encoded in Biological Expression Language**. *Database*, 2018. doi:10.1093/database/bay126
62. **BEL2ABM: agent-based simulation of static models in Biological Expression Language**. *Bioinformatics*, 2018. doi:10.1093/bioinformatics/bty107
63. †**PyBEL: a Computational Framework for Biological Expression Language**. *Bioinformatics*, 2018. doi:10.1093/bioinformatics/btx660
64. †**A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities**. *Database*, 2018. doi:10.1093/database/bay050
65. **A Case Study on the Parametric Occurrence of Multiple Steady States**. *ISSAC 2017*. doi:10.1145/3087604.3087622
66. **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, 2021)
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. **Bioregistry Workshop at Biocuration 2025**. *Bioregistry Workshop @ Biocuration 2025* (April 6, 2025)
2. **Bioregistry and the NFDI in 2024**. *3rd Ontologies4Chem Workshop* (December 9, 2024)
3. **Assembly and Reasoning over Semantic Mappings at Scale**. *Biocuration 2024* (March 7, 2024)
4. **Introduction to WPCI 2023**. *Winter 2023 Workshop on Prefixes, CURIEs, and IRIs* (November 27, 2023)
5. **Standardization of chemical prefixes, CURIEs, URIs, and semantic mappings**. *Ontologies4Chem Workshop 2023* (October 12, 2023)
6. **Improving the reproducibility of cheminformatics workflows with chembl-downloader**. *RDKit User Group Meeting 2023* (September 21, 2023)
7. **Promoting the longevity of curated scientific resources through open code, open data, and public infrastructure**. *Biocuration 2023* (April 26, 2023)
8. **Using dashboards to monitor ontology standardisation and community activity**. *Ontology Summit 2023* (February 15, 2023)
9. **Introduction to WPCI 2022**. *2022 Workshop on Prefixes, CURIEs, and IRIs* (December 5, 2022)
10. **The Bioregistry, CURIEs, and OBO Community Health**. *International Conference on Biomedical Ontology (ICBO)* (September 26, 2022)
11. **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)
12. **A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs**. *Graph Learning Benchmarks (GLB 2022)* (April 26, 2022)
13. **Introduction to WPCI 2021**. *2021 Workshop on Prefixes, CURIEs, and IRIs* (October 29, 2021)
14. **Biomappings: Community Curation of Mappings between Biomedical Entities**. *4th Session of the International Society of Biocuration 2021 Virtual Conference* (October 5, 2021; poster)
15. **The Bioregistry: A Metaregistry for Biomedical Entities**. *12th International Conference on Biomedical Ontologies* (September 17, 2021)
16. **Future Directions for WikiPathway Meta-curation**. *WikiPathways Developers Conference Call* (January 6, 2021)
17. **The Biological Expression Language and PyBEL in 2020**. *COVID-19 Disease Map Community Meeting* (July 10, 2020)
18. **Maintenance and Enrichment of Disease Maps in Biological Expression Language**. *4th Disease Maps Community Meeting* (October 4, 2019; poster)
19. **Identifying Drug Repositioning Candidates using Representation Learning on Heterogeneous Networks**.

The Eighth Joint Sheffield Conference on Chemoinformatics (June 19, 2019; poster)

20. **The PyBEL Ecosystem in 2018**. *OpenBEL Community Meeting* (May 14, 2018)
21. **From Knowledge Assembly to Hypothesis Generation**. *Bio-IT World* (April 22, 2018)
22. **Knowledge Assembly in Systems and Networks Biology**. *Bio-IT World* (April 23, 2018; poster)
23. **The Human Brain Pharmacome: An Overview**. *3rd European Conference on Translational Bioinformatics* (April 17, 2018; poster)
24. **Gene Set Analysis using Phenotypic Screening Data**. *Research, Innovation and Scholarship Expo 2015* (April 9, 2015; poster)