







# Charles Tapley Hoyt

cthoit.com ·  linkedin.com/in/cthoit ·  github.com/cthoit ·  cthoyt ·  Q47475003 ·  https://orcid.org/0000-0003-4423-4370 ·  PjrpzUIAAAAJ

## Fields of Work

Bioinformatics, Pathway Analysis, Machine Learning, Natural Language Processing, Ontology, Knowledge Graph, Proteochemometrics, Drug Repositioning, Systems Biology, Cheminformatics

## Spoken Languages

English, German

## Education

- 2018-19 **Doctor Of Philosophy**, University of Bonn, Germany.
- 2015-17 **Master Of Science**, University of Bonn, Germany.
- 2011-15 **Bachelor Of Science**, Northeastern University, United States of America.
- 2007-11 **High School Diploma**, North Haven High School, United States of America.

## Work

- 2021- **Research Fellow**, *Harvard Medical School*, Bonn, Germany.
- 2020 **Computational Biologist**, *Enveda Biosciences*, Bonn, Germany.
- 2018-19 **Lecturer**, *University of Bonn*, Bonn, Germany.
- 2016-19 **Research Fellow**, *Fraunhofer Institute for Algorithms and Scientific Computing*, Sankt Augustin, Germany.
- 2014 **Intern**, *Novartis*, Cambridge, United States of America.
- 2013-14 **Intern**, *Pfizer*, Cambridge, United States of America.
- 2013 **Intern**, *GlaxoSmithKline*, Waltham, United States of America.
- 2012-15 **Teaching Assistant**, *Northeastern University*, Boston, United States of America.

# Research Output

## Software

**SSSOM-py**, *Python package*.

<https://github.com/mapping-commons/sssom-py/>, <https://mapping-commons.github.io/sssom-py/>

**PyBEL**, *Python library for Biological Expression Language*.

<https://github.com/pybel/pybel>, <https://pybel.github.io>

**Integrated Network and Dynamical Reasoning Assembler**, *Text mining and systems biology modeling software*.

<https://github.com/sorgerlab/indra>, <https://www.indra.bio/>

**PyKEEN**, *Learning, evaluation and applications of knowledge graph embeddings..*

<https://github.com/pykeen/pykeen>, <https://pykeen.github.io>

## Articles 2022

1. **Unifying the Identification of Biomedical Entities with the Bioregistry**

*bioRxiv*, 2022

doi:10.1101/2022.07.08.499378 biorxiv:2022.07.08.499378

2. **Ontology Development Kit: a toolkit for building, maintaining, and standardising biomedical ontologies**  
*arXiv*, 2022  
arxiv:2207.02056
3. **Integrating multi-omics data reveals function and therapeutic potential of deubiquitinating enzymes**  
*eLife*, 2022  
doi:10.7554/elife.72879
4. **Understanding the performance of knowledge graph embeddings in drug discovery**  
*Artificial Intelligence in the Life Sciences*, 2022  
doi:10.1016/j.ailsci.2022.100036
5. **A Simple Standard for Sharing Ontological Mappings (SSSOM)**  
*Database*, 2022  
doi:10.1093/database/baac035
6. **ProtSTonKGs: A Sophisticated Transformer Trained on Protein Sequences, Text, and Knowledge Graphs**  
*SWAT4HCLS 2022*  
ceur-ws:3127:13
7. **Do-calculus enables estimation of causal effects in partially observed biomolecular pathways**  
*Bioinformatics*, 2022  
doi:10.1093/bioinformatics/btac251
8. **A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs**  
*arXiv*, 2022  
arxiv:2203.07544
9. **PyBioPAX: biological pathway exchange in Python**  
*JOSS*, 2022  
doi:10.21105/joss.04136
10. **An Open Challenge for Inductive Link Prediction on Knowledge Graphs**  
*arXiv*, 2022  
arxiv:2203.01520
11. **ChemicalX: A Deep Learning Library for Drug Pair Scoring**  
*arXiv*, 2022  
arxiv:2202.05240
12. **STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs**  
*Bioinformatics*, 2022  
PMID:34986221 PMC8896635 doi:10.1093/bioinformatics/btac001

---

## Articles 2021

13. **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

14. **Gilda: biomedical entity text normalization with machine-learned disambiguation as a service**  
*bioRxiv*, 2021  
doi:10.1101/2021.09.10.459803 biorxiv:2021.09.10.459803
15. **The role of metadata in reproducible computational research**  
*Patterns*, 2021  
PMID:34553169 PMC8441584 doi:10.1016/j.patter.2021.100322
16. **Wavelet-Packet Powered Deepfake Image Detection**  
*arXiv*, 2021  
arxiv:2106.09369
17. **A Review of Biomedical Datasets Relating to Drug Discovery: A Knowledge Graph Perspective**  
*arXiv*, 2021  
arxiv:2102.10062
18. **A Systems Biology Approach for Hypothesizing the Effect of Genetic Variants on Neuroimaging Features in Alzheimer's Disease**  
*JAD*, 2021  
PMID:33554913 PMC8075382 doi:10.3233/jad-201397
19. **PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings**  
*JMLR*, 2021  
arxiv:2007.14175
20. **Leveraging Structured Biological Knowledge for Counterfactual Inference: A Case Study of Viral Pathogenesis**  
*IEEE TBData*, 2021  
doi:10.1109/tbdata.2021.3050680

## Articles 2020

21. **CLEP: A Hybrid Data- and Knowledge- Driven Framework for Generating Patient Representations**  
*bioRxiv*, 2020  
doi:10.1101/2020.08.20.259226 biorxiv:2020.08.20.259226
22. **Extension of Roles in the ChEBI Ontology**  
*ChemRxiv*, 2020  
doi:10.26434/chemrxiv.12591221
23. **The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST)**  
*Bioinformatics*, 2020  
PMID:32637990 doi:10.1093/bioinformatics/btaa622
24. **GuiltyTargets: Prioritization of Novel Therapeutic Targets with Deep Network Representation Learning**  
*IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020  
PMID:32750869 doi:10.1109/tcbb.2020.3003830
25. **PS4DR: a multimodal workflow for identification and prioritization of drugs based on pathway signatures**  
*BMC Bioinf.*, 2020  
PMID:32503412 PMC7275349 doi:10.1186/s12859-020-03568-5
26. **A Computational Approach for Mapping Heme Biology in the Context of Hemolytic Disorders**  
*Frontiers in Bioengineering and Biotechnology*, 2020

## Articles 2019

27. **The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling**  
*Frontiers in Genetics*, 2019  
PMID:31824580 PMC6883970 doi:10.3389/fgene.2019.01203
28. **Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm**  
*BMC Bioinf.*, 2019  
PMID:31604427 PMC6788110 doi:10.1186/s12859-019-3101-1
29. **The KEEN Universe**  
*ISWC 2019*  
doi:10.1007/978-3-030-30796-7<sub>1</sub>
30. **Predicting Missing Links Using PyKEEN**  
*ISWC 2019*  
ceur-ws:2456:64
31. **RatVec: A General Approach for Low-dimensional Distributed Vector Representations via Domain-specific Rational Kernels**  
*LWDA 2019*
32. **BioKEEN: a library for learning and evaluating biological knowledge graph embeddings**  
*Bioinformatics*, 2019  
PMID:30768158 doi:10.1093/bioinformatics/btz117
33. **PathMe: merging and exploring mechanistic pathway knowledge**  
*BMC Bioinf.*, 2019  
PMID:31092193 PMC6521546 doi:10.1186/s12859-019-2863-9
34. **Integration of Structured Biological Data Sources using Biological Expression Language**  
*bioRxiv*, 2019  
doi:10.1101/631812 biorxiv:631812
35. **ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases**  
*NPJ Syst Biol Appl.*, 2019  
PMID:30564458 PMC6292919 doi:10.1038/s41540-018-0078-8
36. **Re-curation and rational enrichment of knowledge graphs in Biological Expression Language Database**, 2019  
PMID:31225582 PMC6587072 doi:10.1093/database/baz068
37. **Challenges of Integrative Disease Modeling in Alzheimer's Disease**  
*Frontiers in molecular biosciences*, 2019  
PMID:31993440 PMC6971060 doi:10.3389/fmolb.2019.00158

## Articles 2018

38. **BEL Commons: an environment for exploration and analysis of networks encoded in Biological Expression Language Database**, 2018  
PMID:30576488 PMC6301338 doi:10.1093/database/bay126

39. **BEL2ABM: agent-based simulation of static models in Biological Expression Language**  
*Bioinformatics*, 2018  
PMID:29949955 PMC6022644 doi:10.1093/bioinformatics/bty107
40. **PyBEL: a Computational Framework for Biological Expression Language**  
*Bioinformatics*, 2018  
PMID:29048466 PMC5860616 doi:10.1093/bioinformatics/btx660
41. **A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities**  
*Database*, 2018  
PMID:29873705 PMC6007221 doi:10.1093/database/bay050

## Articles 2017

42. **A Case Study on the Parametric Occurrence of Multiple Steady States**  
*ISSAC 2017*  
doi:10.1145/3087604.3087622

## Articles 2014

43. **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  
PMID:25127163 PMC4155488 doi:10.1016/j.bmcl.2014.07.063

## Reviews

Note that this section is automatically generated from Wikidata and is not necessarily complete.

1. **The LOTUS initiative for open knowledge management in natural products research**  
*eLife*, 2022  
PMID:35616633 doi:10.7554/elife.70780
2. **New reasons for biologists to write with a formal language**  
*Database*, 2022  
doi:10.1093/database/baac039
3. **PecanPy: a fast, efficient and parallelized Python implementation of node2vec**  
*Bioinformatics*, 2021  
PMID:33760066 doi:10.1093/bioinformatics/btab202

## Acknowledgements

Note that this section is automatically generated from Wikidata and is not necessarily complete.


1. **Automated assembly of molecular mechanisms at scale from text mining and curated databases**  
*bioRxiv*, 2022  
doi:10.1101/2022.08.30.505688 biorxiv:2022.08.30.505688
2. **PecanPy: a fast, efficient and parallelized Python implementation of node2vec**  
*Bioinformatics*, 2021  
PMID:33760066 doi:10.1093/bioinformatics/btab202
3. **The status of causality in biological databases: data resources and data retrieval possibilities to support logical modeling**  
*Briefings in Bioinformatics*, 2020  
PMID:33378765 doi:10.1093/bib/bbaa390

#### 4. **Multimodal Mechanistic Signatures for Neurodegenerative Diseases (NeuroMMSig): a web server for mechanism enrichment**

*Bioinformatics*, 2017

PMID:28651363 doi:10.1093/bioinformatics/btx399

### Mentees

1. Aman Choudhri  0000-0003-4963-6651
  - Student Research Assistant (June - October 2020)
2. Lauren Nicole DeLong  0000-0001-6618-3370
  - Student Research Assistant (September - December 2019)
3. Vinay Bharadhwaj  0000-0002-9901-9494
  - Student Research Assistant (July - December 2019)
4. Mauricio Pio de Lacerda  0000-0002-1770-3007
  - Master's Student (March - December 2019)
5. Yojana Gadiya  0000-0002-7683-0452
  - Student Research Assistant (April - May 2019)
6. Trusha Adeshara  0000-0002-8929-4724
  - Student Research Assistant (April - May 2019)
7. Rana Aldisi  0000-0002-3034-9970
  - Master's Student (March - December 2019)
  - Student Research Assistant (July 2018 - March 2019)
8. Lingling Xu  0000-0002-0303-8616
  - Master's Student (March - December 2019)
  - Student Research Assistant (July 2018 - March 2019)
9. Özlem Muslu  0000-0003-0408-6190
  - Master's Student (May - December 2018)
10. Kristian Kolpeja  0000-0001-9661-5277
  - Student Research Assistant (July - November 2018)
11. Esther Wollert  0000-0002-7128-929X
  - Student Research Assistant (July 2018 - August 2019)
12. Sandra Spalek  0000-0002-6117-4413
  - Student Research Assistant (July 2018 - August 2019)
13. Keerthika Lohanadan
  - Student Research Assistant (July - September 2018)
14. Colin Birkenbihl  0000-0002-7212-7700
  - Student Research Assistant (July - October 2017)
15. Aram Grigoryan
  - Student Research Assistant (July - December 2017)

### Courses

1. Mechanism Enrichment Using Neurommsig (Practical; Winter 2020-2021)  
University of Bonn
2. Mechanism Enrichment Using Neurommsig (Practical; Winter 2019-2020)

University of Bonn

3. Mathematics Meets Life Sciences (Lecture; Winter 2019-2020)  
University of Bonn
4. Enzyme Technology Internship (Practical; Summer 2019)  
University of Bonn
5. Life Sciences Knowledge Discovery (Lecture; Summer 2019)  
University of Bonn
6. Knowledge Assembly, Data Integration, and Modeling in Systems and Networks Biology (Seminar; Winter 2018-2019)  
University of Bonn
7. Biological Databases (Lecture; Winter 2018-2019)  
University of Bonn
8. Life Sciences Knowledge Discovery (Lecture; Summer 2018)  
University of Bonn
9. Biological Databases (Lecture; Winter 2017-2018)  
University of Bonn
10. Life Sciences Knowledge Discovery (Lecture; Summer 2017)  
University of Bonn
11. Biomedical Database Lab (Practical; Winter 2016-2017)  
University of Bonn
12. Drug Discovery and Development (Lecture; Summer II 2015)  
Northeastern University
13. Organic Chemistry II for Majors (Lecture; Spring 2015)  
Northeastern University
14. Organic Chemistry I for Majors (Lecture; Fall 2014)  
Northeastern University
15. Organic Chemistry II for Majors (Lecture; Spring 2014)  
Northeastern University
16. Organic Chemistry I for Majors (Lecture; Fall 2013)  
Northeastern University