Heming Zhang

hemingzhang@wustl.edu heming-zhang.github.io Google Scholar GitHub

Education

Washington University in St. Louis — Ph.D. Candidate, Biomedical Informatics & Data Science — **Aug 2021 – Present**
Multi-Omics Analysis, Graph Neural Network, Large Reasoning Model, Agentic AI, Advised by Prof. Michael Province and Prof. Fuhai Li — *St. Louis, MO*

Washington University in St. Louis — M.S., Computer Science — **Aug 2019 – May 2021**
Drug Discovery, Explainable AI, Graph Neural Network, Co-advised by Prof. Yixin Chen and Prof. Fuhai Li — *St. Louis, MO*

University of British Columbia — Visiting International Research Student (VIRS), Computer Science — **June 2018 - Oct 2018**
Funded by Mitacs Globalink (Top 0.5%), Decentralized Machine Learning, Advised by Prof. Ivan Beschastnikh — *Vancouver, BC*

Central China Normal University — B.S., Information Systems — **Sept 2015 – Jun 2019**
Ranked 1st Among 44, Boya Plan Student (Top 1.5%) — *Wuhan, China*

Papers

- **GALAX: Graph-Augmented Language Model for Explainable Reinforcement-Guided Subgraph Reasoning in Precision Medicine**
  Heming Zhang, Di Huang, Wenyu Li, Michael Province, Yixin Chen, Philip Payne, Fuhai Li
  Preprint, Sept 2025
- **BioMedGraphica: An All-in-One Platform for Biomedical Prior Knowledge and Omic Signaling Graph Generation**
  Heming Zhang*, Shunning Liang*, Tim Xu*, Wenyu Li, Di Huang, Yuhan Dong, etc., Carlos Cruchaga, Yixin Chen, Michael Province, Philip Payne, Fuhai Li
  Preprint, Sept 2025
- **OmniCellAgent: Towards AI Co-Scientists for Scientific Discovery in Precision Medicine**
  Di Huang*, Hao Li*, Wenyu Li*, **Heming Zhang***, etc., Carlos Cruchaga, Michael Province, Yixin Chen, Philip Payne, Fuhai Li
  Preprint, Aug 2025
- **mosGraphFlow: a novel integrative graph AI model mining disease targets from multi-omic data**
  Heming Zhang, Dekang Cao, Tim Xu, Emily Chen, Guangfu Li, Yixin Chen, Philip Payne, Michael Province, Fuhai Li
  Accepted to BMC Methods (in proofreading), Aug 2025
- **OmniCellTOSG: The First Cell Text-Omic Signaling Graphs Dataset for Joint LLM and GNN Modeling**
  Heming Zhang*, Tim Xu*, Dekang Cao*, Shunning Liang, Lars Schimmelpfennig, Levi Kaster, Di Huang, Carlos Cruchaga, etc., Philip Payne, Fuhai Li
  Preprint, April 2025
- **M3NetFlow: a novel multi-scale multi-hop modular graph AI model for multi-omics data integration and signaling network inference**
  Heming Zhang, Peter Goedegebuure, Li Ding, William Hawkins, David DeNardo, Ryan Fields, Yixin Chen, Fuhai Li
  iScience, Feb 2025
- **GraphSeqLM: A Unified Graph Language Framework for Omic Graph Learning**
  Heming Zhang, Di Huang, Yixin Chen, Fuhai Li
  WWW 2025
- **Using DeepSignalingFlow to mine signaling flows interpreting mechanism of synergy of cocktails**
  Heming Zhang, Yixin Chen, Philip R Payne, Fuhai Li
  NPJ Systems Biology and Applications, Aug 2024
- **mosGraphGen: a novel tool to generate multi-omic signaling graphs to facilitate integrative and interpretable graph AI model development**
  Heming Zhang*, Dekang Cao*, Zirui Chen, Ziyuan Zhang, Yixin Chen, Cole Sessions, etc., Guangfu Li, Michael Province, Fuhai Li
  Bioinformatics Advances, Sept 2024
- **Interpreting the Mechanism of Synergism for Drug Combinations Using Attention-Based Hierarchical Graph Pooling**
  Zehao Dong, **Heming Zhang**, Yixin Chen, Philip R Payne, Fuhai Li
  Cancers, Aug 2023
- **Predicting anti-cancer drug response with deep learning constrained by signaling pathways**
  Heming Zhang, Yixin Chen, Fuhai Li
  Frontiers in Bioinformatics, Mar 2021
- **Investigate the relevance of major signaling pathways in cancer survival using a biologically meaningful deep learning model**
  Jiarui Feng, **Heming Zhang**, Fuhai Li
  BMC Bioinformatics, Oct 2020
- **Predicting Tumor Cell Response to Synergistic Drug Combinations Using a Novel Simplified Deep Learning Model**
  Heming Zhang, Jiarui Feng, Amanda Zeng, Philip Payne, Fuhai Li
  AMIA 2020 (Oral Presentation)

Projects

Develop an AI Co-Scientist for Scientific Reasoning in Biomedicine — **April 2025 – Present**
Funded by NLM R01 LM013902; NIA R21 AG078799; NIA R56 AG065352; NINDS RM1 NS132962
- Design and improve **OmniCellAgent**—built on **BioMedGraphica** and **Graph-Language Foundation Models (GLFMs)**—integrating data querying, GLFM inference, and explanation generation in a closed loop to produce interpretable hypotheses (e.g., target prioritization, drug-combination suggestions).
- Finetune the agent orchestrator to enhance task planning using reinforcement learning with reward modeling.
- Establish a benchmark using expert knowledge, literature alignment, and human feedback to evaluate scientific reasoning in biomedical agentic AI.

Biomedical AI Ecosystem for Multi-Omics Integration and Analysis — **Feb 2024 – Present**
Funded by NLM R01 LM013902; NIA R21 AG078799; NIA R56 AG065352; NINDS RM1 NS132962
- Preprocess DNA methylation data based on CpG sites, then fuse multi-omics (epigenomic/genomic/transcriptomic/proteomic) into a signaling graph by **mosGraphGen** to generate a graph-AI ready dataset.
- Build **BioMedGraphica**, an all-in-one platform that harmonizes biomedical nomenclature across fragmented resources and, from user-specified inputs, auto-generates a **Text–Numeric Graph (TNG)**—a novel data format bridging textual biological priors with quantitative features—to develop **Graph–Language Foundation Models (GLFMs)**. To facilitate broad adoption, the platform is released as an **public web interface**.
- Release two ontology-standardized, distribution-calibrated **TNG** datasets: (1) **OmniCellTOSG** for single-cell data (CellxGene, GEO, BrainCellAtlas) and

(2) **MOTASG** for bulk omics data (TCGA, DepMap, Synapse), enabling GLFMs pretraining and benchmarking downstream tasks for disease classification, survival prediction, CRISPR essentiality, drug response, and cell-type annotation.
- Enable interpretability via attention-based attribution or reinforcement-guided subgraph reasoning for mechanistic insights.

AI for Aging and Longevity: Integrating Omics and Phenotypic Data **June 2023 – Present**
Funded by NIA R56 AG065352; research using the Long Life Family Study (LLFS) cohort
- Curate and preprocess **LLFS** records (**1,405** patients; **122** sub-features across **41** groups) and stratify labels into **187** pre-T2D, **123** T2D, and **1,095** no-T2D.
- Introduce a novel graph AI framework **Graph in Graph (GiG)** to embed omics signaling graphs into a person–phenotype graph to fuse clinical, demographic, and molecular modalities while handling sparsity and mixed data types. Meanwhile, we integrate **GWAS** information by projecting variant-level signals to genes and fusing them into the signaling graph.
- Provide interpretability via attention weights and pooled-subgraph rationales, highlighting key clinical features, signaling genes, and pathways.

Graph AI for Precision Medicine in Lower Urinary Tract Symptoms (LUTS) **Oct 2021 – Sept 2023**
Supported by NIDDK U01 DK100017-09; research based on the Lower Urinary tract dysfunction Research Network (LURN) cohort
- Curate and preprocess LURN cohort (**1,053** patients, **208** sub-features across **60** characteristic groups) spanning demographics, surveys, treatments, and comorbidities. Also, we used 2 standards to split overactive bladder (OAB) medication responders, defining response labels with medication-timing rules: Level 1 (87 responders / 47 non-responders) and Level 2 (65 responders/ 69 non-responders).
- Build **LUTSPheNet**, a graph AI framework that models patient–phenotype relations to predict OAB medication response in LUTS, achieving best predictive performance than baseline models.

Pathway-Constrained Deep Learning for Precision Oncology and Drug Discovery **Feb 2020 – Aug 2024**
Funded by Children's Discovery Institute (CDI) M-II-2019-802; startup support from I2DB and the Department of Pediatrics, WashU
- Build pathway-constrained models that link genes to **46 cancer signaling pathways**, using **CCLE** with **GDSC** responses, **NCI ALMANAC** drug-combination screens, and **TCGA** multi-omics plus clinical data.
- Predict **single-drug response**, **drug-combination synergy**, and **patient survival** with biologically grounded, interpretable outputs at the pathway level.
- Achieve consistent gains over vanilla DNN baselines while highlighting candidate mechanisms and targets to guide precision oncology.

Professional Experience

Research Assistant **Feb 2020 – Present**
Institute for Informatics, Data Science and Biostatistics, Washington University School of Medicine *St. Louis, MO*
- Build an AI Co-Scientist for Autonomous Scientific Reasoning in Biomedicine
- Develop biomedical AI ecosystem for interpretable multi-omics integration and discovery

Teaching Assistant **Jan 2023 – May 2023**
BMDS 5305: Introduction to Biomedical Data Science II, Washington University School of Medicine *St. Louis, MO*
- Delivered two hands-on lectures covering basic neural networks, convolutional neural networks, and graph neural networks.
- Held weekly TA hours; guided students on assignments and final projects.

Teaching Assistant **Jan 2020 – May 2020**
CSE 417T: Introduction to Machine Learning, Department of Computer Science & Engineering, Washington University in St. Louis *St. Louis, MO*
- Ran office hours to clarify lecture content and problem sets.
- Supported implementations of PLA, logistic regression, bagged trees/random forest, and AdaBoost.

Visiting International Research Student (VIRS) **Jul 2018 – Oct 2018**
Department of Computer Science & Center for Decision-Making and Action, University of British Columbia *Vancouver, BC*
- Awarded **Mitacs Globalink Research Internship** ($4,500).
- Prototyped a privacy-preserving, decentralized blockchain workflow for distributed machine learning.

Academic Service

Journal Reviewer
- Briefings in Bioinformatics
- Scientific Reports
- BMC Bioinformatics
- BMC Genomics
- NPJ Systems Biology and Applications
- Journal of Cheminformatics
- Discover Oncology
- Journal of Translational Medicine

Invited Talks
- International World Wide Web Conference 2025, Sydney, Australia
- 12th International Conference on Intelligent Biology and Medicine (ICIBM 2024), Huston, TX
- AMIA 2020 Virtual Annual Symposium, Chicago, IL (Transferred to virtual due to COVID-19)

Awards and Honors

Awards
- ICIBM 2024 Travel Award with (2024, $600)
- Mitcas Global Link Scholarship (2018, $4,500)

Honors
- Outstanding Graduate, CCNU (2019, Top 20%)
- Grand Prize, Hubei Challenge Cup (2017, Top 3%)
- Boya Plan Student, CCNU (2016, Top 1.5%)