Michael Cheng

Portfolio: https://mcheng7777.github.io | Email: mcheng7777@g.ucla.edu | Phone: (510)-648-1238
LinkedIn: https://www.linkedin.com/in/mcheng7777/ | GitHub: https://github.com/mcheng7777

SUMMARY

A Ph.D. candidate in Bioinformatics specializing in agentic systems, multi-omics machine learning, drug discovery, and precision medicine. Seeking a machine learning developer role to create innovative AI-driven tools for therapeutic development and biomedical research.

SKILLS

Programming & Databases: Python, HTML, CSS, JavaScript, R, C++, Bash, SQL, Cypher, Neo4j, CUDA
Machine Learning and Data Science: PyTorch, HuggingFace, Langchain/Langgraph/Langsmith, Pandas, NumPy, Matplotlib, Seaborn, Scanpy, CuPy, CuML
AI Tools: Langchain, Langgraph, Langsmith, PydanticAI, RAG/GraphRAG, FastMCP
Bioinformatics Concepts: Graph Neural Networks, Single Cell and Spatial Genomics, GWAS, Precision Medicine
Web Development: FastAPI, React, Next.js, Streamlit, Plotly/Dash, JavaScript, JQuery/AJAX, HTML, CSS, PHP
Cloud and DevOps: Google Cloud Platform (GCP), Docker, Git, High Performance Computing, GitHub actions

PROJECTS

NVIDIA x Scverse Hackathon: GPU-accelerated single cell genomics analysis. Achieved > 100x speedups against current single cell libraries and actively contributing to the RAPIDS-singlecell GitHub.
PubmedRAG: A literature search RAG solution to accelerate introductory research for a user by curating PubMed articles related to the requested key words and articles.
Biorag-lab: Reinforcement learning-based coding assistant and dataset finder to reproduce publication code and test on external datasets.

RESEARCH EXPERIENCE

Bioinformatics AI Intern | Refined Science | Denver, CO (Remote)
June 2025 – Present

- **AI Systems**: Built a research report agent to generate insightful reports based on up-to-date biomedical knowledge graphs and drug market intelligence using GraphRAG and SQLRAG.
- **Evaluation Frameworks**: Developed LLM-as-judge evaluation framework for accuracy and retrieval relevance.
- **Rapid Prototyping**: Developed chatbot UIs in Dash and Streamlit featuring reasoning steps to solicit and incorporate feedback from key opinion leaders.
- **Skills**: Langchain, Langgraph, Langsmith, FastAPI, Streamlit, Dash, Docker, PostrgreSQL, Neo4j, GCP, GitHub

Ph.D. Researcher, Advisor: Professor Xia Yang | UCLA Bioinformatics Department | Los Angeles, CA
April 2022 - Present
Multiomics machine learning for drug target discovery and precision medicine.

- **Machine Learning**: Developed gradient boosting model to infer gene regulation networks from scRNA-seq data with 0.5 AUROC improvement over existing methods in drug effect prediction on gene expression. Published in *iScience*.
- **Graph Neural Networks**: Training graph attention networks on spatial transcriptomics and scRNA-seq data for cell communication inference.
- **Precision Medicine:** Integrated scRNA-seq with genome wide association studies (GWAS) to identify genetic causes of memory deficits in traumatic brain injury and calculated polygenic risk scores to improve patient stratification and diagnosis. Submitted to *NPJ Systems Biology and medRxiv*.
- **Drug Target Discovery:** Integrated scRNA-seq gene networks and drug signature databases to discover and validate 4 drug targets that weaken cancer resistance to immunotherapy. Submitted to *PNAS and bioRxiv*.
- **Web Development & Data Engineering:** Created and deployed the first cell type-specific gene network database and web platform across millions of cells from public scRNA-seq data repositories for network biology research using LAMP stack, Neo4j, and Jenkins. Presented at *RECOMB/ICSB* and *Cell Symposia* Conferences 2023.

Bioinformatics Researcher, Hilary Coller Lab | UCLA Department of Molecular Biology | Los Angeles, CA
September 2020 – October 2023
Multiomics representation learning of cancer heterogeneity. Published in *Communications Biology* and featured by the National Cancer Institute.

- **Representation Learning:** Performed nonnegative matrix factorization on RNA-seq and methylation data of 24 cancer types in The Cancer Genome Atlas to identify latent epigenomic features that correlate with cancer survival.

- **Neural Networks:** Trained neural network to predict log hazard of cancer patients from RNA-seq data, which significantly stratified testing cohort into high and low survival groups.

FEATURED PUBLICATIONS

Cheng, M., Mitra, M. & Coller, H. Pan-cancer landscape of epigenetic factor expression predicts tumor outcome. *Communications Biology*. (2023).

Littman, R., **Cheng, M**., Wang, N., Peng, C. & Yang, X. SCING: Inference of robust, interpretable gene regulatory networks from single cell and spatial transcriptomics. *iScience*. (2023).

Cheng, M., Mao, M., …, & Yang, X. Interactions between mild traumatic brain injury and genetics perturb neuronal and glial pathways and networks relevant to learning and memory in ABCD study. *medRxiv.* (2025).

HONORS AND AWARDS

Dissertation Year Award – UCLA Graduate Division (2025)

- Received $30,000 to support the completion of my dissertation research.

Hyde Fellowship Award – UCLA Department of Integrative Biology and Physiology (2024)

- Received $42,600 to support my dissertation research in multiomics machine learning methods development.

EDUCATION

University of California, Los Angeles (UCLA) | Los Angeles, CA

- *Ph.D., Bioinformatics* | Overall GPA: 4.0/4.0 | Anticipated June 2026
- *B.S., Molecular Biology, Bioinformatics* | Overall GPA: 3.919/4.0 | June 2021