YIMIN ZHAO

SEATTLE, WA | (+1) 410-814-9182 | ymzhao.whu@gmail.com | Google Scholar | LinkedIn

SUMMARY

PhD candidate in Biostatistics specializing in artificial intelligence and its real-world applications to complex, high-dimensional data. Led a rubric-based LLM evaluation pipeline; built a Retrieval-Augmented Generation (RAG) tool for multi-omics analysis recommendations; developed semi-supervised conformal-inference methods for single-cell branch-point localization; built a multimodal VAE model for multi-modality imputation. Collaborate closely with domain experts to deliver trustworthy, scalable AI solutions across biomedical and other data-intensive domains.

SKILLS

ML/NLP	LLM evaluation(Rubrics, Factuality), RAG, Prompt Engineering, Deep Learning, Generative Models Conformal Inference, Statistical Machine Learning, Semi-supervsied Learning, A/B Testing
Software	Python(Pytorch, Numpy, Pandas, LangChain), R, Shell, SQL

EDUCATION

University of Washington, School of Public Health — Seattle, WA
Ph.D. in Biostatistics (*Advisor: Jeff Leek, Mike Wu, Kevin Lin*) — Sep. 2022 - May. 2027

Johns Hopkins University, Bloomberg School of Public Health — Baltimore, MD
ScM in Biostatistics (*Advisor: Hongkai Ji*) — Sep. 2020 - May. 2022

Wuhan University, School of Mathematics and Statistics — Wuhan, China
B.S. in Mathematics and Applied Mathematics — Sep. 2015 - Sep. 2019

PUBLICATIONS

Yimin Zhao, Connor Finkbeiner, Manu Setty, Kevin Z. Lin(2026). SCOPE: Localizing fate-decision states and their regulatory drivers in single-cell differentiation. *Under review at Nature Communications*

Yimin Zhao, Sheela R Damle, Simone E Dekker, Scott Geng, Karly Williams Silva, Jesse J Hubbard, Manuel F Fernandez, Fatima Zelada-Arenas, Alejandra Alvarez, Brianne Flores, Alexis Rodriguez, Stephen Salerno, Carrie Wright, Zihao Wang, Pang Wei Koh, Jeffrey T Leek(2026). PanCanBench: A Comprehensive Benchmark for Evaluating Large Language Models in Pancreatic Oncology. *Under review at NEJM AI*

Zhen Zeng, Tianbei Zhang, Jiajia Zhang, Shuai Li, Sydney Connor, Boyang Zhang, **Yimin Zhao**, Jordan Wilson, Dipika Singh, Rima Kulikauskas, Candice D. Church, Thomas H. Pulliam, Saumya Jani, Paul Nghiem, Suzanne L. Topalian, Patrick M. Forde, Drew M. Pardoll, Hongkai Ji, Kellie N. Smith(2025). A minimal gene set characterizes multiple classes of tumor-specific TIL among different cancer types. *Nature Communications*, 16(1): 1070

Zhanying Feng, Xianwen Ren, Yuan Fang, Yining Yin, Chutian Huang, **Yimin Zhao** and Yong Wang(2020). scTIM: Seeking Cell-Type-Indicative Marker from single cell RNA-seq data by optimization model. *Bioinformatics*, 36(8), 2474-2485

RESEARCH PROJECTS

Rubric-Based Evaluation of LLMs for Pancreatic Cancer Q&A — Apr 2025 – Feb 2026
Supervisor: Jeff Leek, Fred Hutch; Pang Wei Koh, UW CSE — Seattle, USA

- Developed an end-to-end pipeline to collect, standardize, and polish expert-rubric annotations from medical fellows for evaluating LLM responses to patient-focused pancreatic cancer questions.
- Created a high-quality dataset of real-world patient questions paired with expert-annotated rubrics.
- Benchmarked commercial (GPT, Claude, Gemini, Grok) and open-source (Llama, Qwen, Olmo) models on the rubric-annotated dataset using an LLM-as-a-judge evaluation framework.

Multi-omics Data Analysis Recommendation with RAG — Apr 2025 – Present
Supervisor: Mike Wu, Fred Hutch — Seattle, WA

- Developed a RAG-based tool for multi-omics analysis recommendation that mitigates hallucinations and delivers concrete, citation-backed guidelines compared with state-of-the-art LLM baselines.

Branchpoint Localization in Development for Single-cell Sequencing Data — Sep 2023 - Mar 2026
Supervisor: Kevin Lin, University of Washington — Seattle, USA

- Developed a computational method to localize branch point in cell development based on conformal inference, semi-supervised learning and density clustering.