







# Qiao Liu

---

CONTACT INFORMATION      300 Geroge Street, Suite 501  
New Haven, CT, 06511  
Tel: +1 203-737-5110  
Email: qiao.liu@yale.edu  
     

RESEARCH INTEREST      Generative AI; Gene Regulation; Causal Inference; Multiomics Integration;  
Single Cell Genomics; Pharmacogenomics; Genomic Foundation Models

PROFESSIONAL EXPERIENCE      **Assistant Professor**      Sep 2025 - Present  
Department of Biostatistics      New Haven, CT  
Yale University

**Postdoctoral Scholar**      Jul 2021 - Aug 2025  
Department of Statistics      Stanford, CA  
Advisor: Prof. Wing Hung Wong

EDUCATION      **Stanford University**      Aug 2019 - Jul 2021  
Visiting Ph.D., Department of Statistics      Stanford, CA  
Advisor: Prof. Wing Hung Wong

**Tsinghua University**      Sep 2016 - Aug 2019  
Ph.D. student, Department of Automation      Beijing, China  
Advisor: Prof. Rui Jiang

**Beihang University**      Sep 2016 - Aug 2019  
B.E., ShenYuan Honors College      Beijing, China

MANUSCRIPTS SUBMITTED      (†=co-first author; \*=corresponding author)

1. Erpai Luo, Lei Wei, Minsheng Hao, Xuegong Zhang\*, Qiao Liu\* Multi-modal diffusion model with dual-cross-attention for multi-omics data generation and translation[J]. bioRxiv, 2025 (in revision at Nature Communications)
2. Qiao Liu and Wing Hung Wong. An AI-powered Bayesian generative modeling approach for causal inference in observational studies. arXiv preprint arXiv:2501.00755, 2025. (in revision at Journal of the American Statistical Association)
3. Qiao Liu†, Wanwen Zeng†, Hongtu Zhu, Lexin Li\*, Wing Hung Wong\*. Leveraging genomic large language models to enhance causal genotype-brain-clinical pathways in Alzheimer’s disease[J]. medRxiv, 2025.

4. Wanwen Zeng<sup>†</sup>, Shuang Chen<sup>†</sup>, Yuti Liu<sup>†</sup>, Qiao Liu, and Wing Hung Wong. bpBERT: an interpretable BERT-based model of base-pair resolution transcription-factor binding reveals DNA sequence regulatory syntax and functional impact of genetic variants[J]. *bioRxiv*, 2025.

PEER-REVIEWED  
PUBLICATIONS

1. Xuejian Cui, Qijin Yin, Zijing Gao, Zhen Li, Xiaoyang Chen, Shengquan Chen, Qiao Liu, Wanwen Zeng\*, Rui Jiang\*. CREATE: cell-type-specific cis-regulatory elements identification via discrete embedding[J]. *Nature Communications*, 2025.
2. Zhaoyang Zhang, Ziqi Chen, Qiao Liu, Jinhan Xie, Hongtu Zhu. Sampling-guided heterogeneous graph neural network with temporal smoothing for scalable longitudinal data imputation. Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining(*KDD*), 2025.
3. Wanwen Zeng, Hanmin Guo, Qiao Liu, Wing Hung Wong. How to improve polygenic prediction from whole-genome sequencing data by leveraging predicted epigenomic features?[J]. *PNAS*, 2025.
4. Wei Zhang, Xianglin Zhang, Qiao Liu, Lei Wei, Xu Qiao, Rui Gao, Zhiping Liu, Xiaowo Wang. Deconer: An evaluation toolkit for reference-based deconvolution methods using gene expression data[J]. *Genomics, Proteomics & Bioinformatics*, 2025.
5. Zijing Gao<sup>†</sup>, Qiao Liu<sup>†\*</sup>, Wanwen Zeng, Wing Hung Wong\*, Rui Jiang\*. EpiGePT: a Pretrained Transformer model for human epigenomics[J]. *Genome Biology*, 2024.
6. Qiao Liu, Zhongren Chen, Wing Hung Wong\*. An encoding generative modeling approach to dimension reduction and covariate adjustment in causal inference with observational studies [J]. Proceedings of the National Academy of Sciences of the United States of America(*PNAS*), 2024, 121(23), e2322376121.
7. Wei Shao, Yuti Liu, Shuang Zhang, Shuang Chen, Qiao Liu, Jianyu Zhou, Wanwen Zeng. Inferring gene regulatory network based on scATAC-seq data with gene perturbation[C]. IEEE International Conference on Bioinformatics and Biomedicine (*BIBM*), 2024.
8. Chuanqi Lao, Pengfei Zheng, Hongyang Chen, Qiao Liu, Feng An, Zhao Li. Comprehensive tissue deconvolution of cell-free DNA by deep learning for disease diagnosis and monitoring[J]. *BMC bioinformatics*, 25 (1), 105, 2024.

9. Shuo Li, Weihua Zeng, Xiaohui Ni, Qiao Liu, Wenyan Li, Mary L. Stackpole, Yonggang Zhou, Arjan Gower, Kostyantyn Krysan, Preeti Ahuja, David S. Lu, Steven S. Raman, William Hsu, Denise R. Aberle, Clara E. Magyar, Samuel W. French, Steven-Huy B. Han, Edward B. Garon, Vatche G. Agopian, Wing Hung Wong\*, Steven M. Dubinett\* and Jasmine Zhou\*. Comprehensive tissue deconvolution of cell-free DNA by deep learning for disease diagnosis and monitoring[J]. Proceedings of the National Academy of Sciences of the United States of America(*PNAS*), 120 (28) e2305236120, 2023.
10. Shuang Zhang, Yuti Liu, Shuang Chen, Qiao Liu, Wanwen Zeng. Applications of Transformer-based language models in bioinformatics: A survey[J]. *Bioinformatics Advances*, 2023.
11. Qiao Liu<sup>†</sup>, Wanwen Zeng<sup>†</sup>, Wei Zhang, Sicheng Wang, Hongyang Chen, Rui Jiang\*, Mu Zhou\*, Shaoting Zhang\*. Deep generative modeling and clustering of single cell Hi-C data[J]. *Briefings in Bioinformatics*, 2023, 24(1): bbac494.
12. Qijin Yin, Xusheng Cao, Rui Fan, Qiao Liu\*, Rui Jiang\* and Wanwen Zeng\*. DeepDrug: A general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction[J]. *Quantitative Biology*, 2023, 11(3), 260-274.
13. Wanwen Zeng<sup>†</sup>, Qiao Liu<sup>†</sup>, Qijin Yin<sup>†</sup>, Rui Jiang\*, Wing Hung Wong\*. HiChIPdb: a comprehensive database of HiChIP regulatory interactions[J]. *Nucleic Acids Research*. 2022.
14. Christopher Lance, Malte D. Luecken, Daniel B. Burkhardt, Robrecht Cannoodt, Pia Rautenstrauch, Anna Laddach, Aidyn Ubungazhibov, Zhi-Jie Cao, Kaiwen Deng, Sumeer Khan, Qiao Liu, Nikolay Russkikh, Gleb Ryazantsev, Uwe Ohler, NeurIPS 2021 Multimodal data integration competition participants, Angela Oliveira Pisco\*, Jonathan Bloom\*, Smita Krishnaswamy\*, Fabian J Theis\*. Multimodal single cell data integration challenge: results and lessons learned[J]. Proceedings of Machine Learning Research(*PMLR*), 2022.
15. Zhana Duren, Fengge Chang, Fnu Naqing, Jingxue Xin, Qiao Liu, Wing Hung Wong\*. Regulatory analysis of single cell multiome gene expression and chromatin accessibility data with scREG[J]. *Genome Biology*, 2022, 23(1): 1-19.
16. Qiao Liu, Kui Hua, Xuegong Zhang, Wing Hung Wong\*, Rui Jiang\*. DeepCAGE: incorporating transcription factors in genome-wide prediction of chromatin accessibility[J]. *Genomics, Proteomics & Bioinformatics*, 2022.

17. Jinxiang Ou, Yunheng Shen, Feng Wang, Qiao Liu, Xuegong Zhang, Hairong Lv\*. AggEnhance: Aggregation enhancement by class interior points in federated learning with non-IID data[J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2022.
18. Qijin Yin, Qiao Liu, Zhuoran Fu, Rui Jiang\*. scGraph: a graph neural network-based approach to automatically identify cell types[J]. *Bioinformatics*, 2022.
19. Tianxing Ma, Qiao Liu, Haochen Li, Mu Zhou, Rui Jiang, Xuegong Zhang\*. DualGCN: a dual graph convolutional network model to predict cancer drug response[J]. *BMC bioinformatics*, 2022, 23(4): 1-13.
20. Qiao Liu, Shengquan Chen, Rui Jiang\*, Wing Hung Wong\*. Simultaneous deep generative modeling and clustering of single cell genomic data[J]. *Nature Machine Intelligence*, 2021, 3(6): 536-544.
21. Qiao Liu, Jiaze Xu, Rui Jiang\*, Wing Hung Wong\*. Density estimation with deep generative neural networks [J]. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 2021, 118(15), e2101344118.
22. Feng Wang, Guoyizhe Wei, Qiao Liu, Jinxiang Ou, Xian Wei, Hairong Lv\*. Boost neural networks by checkpoints [C]. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021, 33.
23. Shengquan Chen, Qiao Liu, Xuejian Cui, Rui Jiang\*. OpenAnnotate: a web server to annotate the chromatin accessibility of genomic regions[J]. *Nucleic Acids Research*, 2021, 49(W1): W483-W490.
24. Qiao Liu, Zhiqiang Hu, Rui Jiang\*, Mu Zhou\*. Cancer drug response prediction via a hybrid graph convolutional network[J]. *Bioinformatics*, 2020. 36(Supplement\_2): i911-i918. Also in *Proceedings of the 19th European Conference on Computational Biology (ECCB)*, 2020.
25. Kexin Ding, Qiao Liu, Edward Lee, Mu Zhou, Aidong Lu, Shaoting Zhang\*. Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer[C]. *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2020, (pp. 294-304).
26. Qingzhu Yang, Qiao Liu, Hairong Lv\*. A decentralized system for medical data management via blockchain [J]. *Journal of Internet Technology*, 2020, 21(5): 1335-1345.

27. Junfeng Liu, Qiao Liu, Qingzhu Yang\*. Mstree: a multispecies coalescent approach for estimating ancestral population size and divergence time during speciation with gene flow [J]. *Genome Biology and Evolution*, 2020, 12(5): 715-719.
28. Chencheng Xu, Qiao Liu, Jianyu Zhou, Minzhu Xie, Jianxing Feng, Tao Jiang\*. Quantifying functional impacts of regulatory variants with multi-task Bayesian neural network[J]. *Bioinformatics*, 2020, 36(5): 1397-1404.
29. Chencheng Xu, Qiao Liu, Minlie Huang, Tao Jiang\*. Reinforced molecular optimization with neighborhood-controlled grammars[C]. Conference on Neural Information Processing Systems (*NeurIPS*), 2020, 33.
30. Qiao Liu, Hairong Lv, Rui Jiang\*. hicGAN infers super resolution Hi-C data with generative adversarial networks. *Bioinformatics*, 2019, 35(14): i99-i107. Also in *Proceedings of the 27th Intelligent Systems for Molecular Biologythe 18th European Conference on Computational Biology (ISMB/ECCB)*, 2019.
31. Qijin Yin, Mengmeng Wu, Qiao Liu, Rui Jiang\*. DeepHistone: a deep learning approach to predicting histone modifications[J]. *BMC Genomics*, 2019,20(2):193.
32. Shaoming Song, Hongfei Cui, Shengquan Chen, Qiao Liu, Rui Jiang\*. Epi-FIT: Functional interpretation of transcription factors based on combination of sequence and epigenetic information[J]. *Quantitative Biology*, 2019, 1-11.
33. Qiao Liu, Fei Xia, Qijin Yin, Rui Jiang\*. Chromatin accessibility prediction via a hybrid deep convolutional neural network[J]. *Bioinformatics*, 2017, 34(5): 732-738.
34. Qiao Liu, Mingxin Gan, Rui Jiang\*. A sequence-based method to predict the impact of regulatory variants using random forest[J]. *BMC Systems Biology*, 2017, 11(2): 7. Also in *Proceedings of the 15th Asia Pacific Bioinformatics Conference (APBC)*, 2017.
35. Bai Li, Mu Lin, Qiao Liu, Ya Li\*, Changjun Zhou\*. Protein folding optimization based on 3D off-lattice model via an improved artificial bee colony algorithm[J]. *Journal of Molecular Modeling*, 2015, 21(10): 261.

HONORS AND  
AWARDS

<b>Pathway to Independence Award (K99/R00)</b> National Institutes of Health (NIH), NHGRI	Aug 2024
<b>ECCB Fellowship</b> International Society for Computational Biology (ISCB)	Aug 2020

	<b>ISMB Travel Award</b> International Society for Computational Biology (ISCB)	Apr 2019
	<b>National Scholarship</b> Ministry of Education of China	Oct 2018
	<b>Microsoft Young Fellowship</b> Microsoft Research Asia	Jun 2015
TALKS AND PRESENTATIONS	<b>Invited Talks and Lectures</b>	
	Duke Boot Camp for AI, Duke University, NC <i>CGM-AI: Causal Generalist Medical AI</i> Jointly with Dr. Hongtu Zhu, Dr. Haoxiu Yao, and Dr. Xin Wang	May 2025
	MCBIOS, University of Utah, UT <i>Multi-modal diffusion model with dual-cross-attention for multi-omics data generation and translation</i>	Mar 2025
	USC QCB Seminar, University of Southern California, CA <i>A Flexible Generative AI Framework for High-dimensional Data Analysis</i>	Nov 2024
	Tsinghua Statistics Seminar, Tsinghua University (remote) <i>CausalEGM: a general causal inference framework by encoding generative modeling</i>	Apr 2023
	JSM, Washington DC <i>scTrace: Time series single-cell trajectory inference using deep generative models</i>	Aug 2022
	CVI Early Career Research Roundtable, Stanford University, CA <i>The application of deep generative models in biology</i>	Jun 2022
	1 <sup>st</sup> place in NeurIPS Multimodal Data Integration, remote talk <i>Single cell multi-modal integrative analysis</i>	Dec 2021
	ECCB, remote talk <i>DeepCDR: a hybrid graph convolutional network for predicting cancer drug response</i>	Sep 2020
	ISMB, Basel, Switzerland <i>hicGAN infers super resolution Hi-C data with generative adversarial networks</i>	Jul 2019
	APBC, Shenzhen, China <i>A sequence-based method to predict the impact of regulatory variants using random forest</i>	Jan 2017

### Contributed Conference Presentations

JSM, Nashville, TN Aug 2025  
*An AI-powered Bayesian generative modeling approach for causal inference in observation studies*

STAGEN, University of Minnesota-Twin Cities, MN May 2025  
*Genomic large language models and the applications in Alzheimer's disease*

ICIBM, Houston, TX October 2024  
*EpiGePT: a pretrained transformer-based language model for context-specific human epigenomics*

NIH CEGS Annual Meeting, Duke University, NC August 2022  
*Analyzing spatially resolved transcriptomics data with deep generative model*

CEGS seminar, Stanford University, CA March 2020  
*Simultaneous deep generative modelling and clustering of single-cell genomic data*

GANDS **K99/R00HG013661**, NIH/NHGRI, PI Sep 2024 - Aug 2028  
Bridging the gap between genetic variants and radiomic phenotypes via genomic large language models  
Total cost: \$1,068,930

TEACHING *STATS 371: Applied Bayesian Statistics* Spring 2025  
Department of Statistics, Stanford University  
Guest Lecturer

*STATS 319: Literature of Statistics* Spring 2024  
Department of Statistics, Stanford University  
Teaching Assistant

*Introduction to Generative Models* Summer 2023  
ICME, Stanford University  
Teaching Assistant

*Intelligent Algorithms and Systems* Fall 2019  
Fundamental Industry Training Center, Tsinghua University  
Teaching Assistant

*Introduction to Artificial Intelligence* Fall 2017  
Department of Automation, Tsinghua University  
Teaching Assistant

SERVICE **Conference Session Chair**

**Journal Reviewer**

*Bioinformatics*  
*Briefings in Bioinformatics*  
*Bioinformatics Advances*  
*BMC Bioinformatics*  
*BMC Genomics*  
*Complex & Intelligent Systems*  
*Computational Statistics*  
*Drug Discovery Today*  
*Engineering Applications of Artificial Intelligence*  
*Genome Biology*  
*Genomics, Proteomics & Bioinformatics*  
*IEEE Transactions on Medical Imaging*  
*iScience*  
*Journal of the American Statistical Association*  
*Journal of Causal Inference*  
*Journal of Computational and Graphical Statistics*  
*Nature Methods*  
*Nature Machine Intelligence*  
*Nature Communications*  
*NAR Genomics and Bioinformatics*  
*Proceedings of the National Academy of Sciences*  
*Patterns*  
*PLOS Computational Biology*

**Conference Reviewer (PC Member)**

International Conference on Learning Representations (ICLR) 2025-26  
Association for the Advancement of Artificial Intelligence (AAAI) 2023-25

PROFESSIONAL  
MEMBERSHIP

American Statistical Association (ASA)  
International Society of Computational Biology (ISCB)  
International Chinese Statistical Association (ICSA)