

# Qiao Liu – Curriculum Vitae

---

<b>Address</b>	300 George Street New Haven, CT 06511	<b>Email</b>	qiao.liu@yale.edu
		<b>Website</b>	https://liuq-lab.com

## Research Statement

I'm an Assistant Professor at the Department of Biostatistics, Yale University. My primary research interest lies in this multi-disciplinary area where I have been devoting to developing AI-powered computational frameworks with significance in both statistical theory and applications. In particular, I have been exploiting the advances in *generative AI* to tackle important high-dimensional statistical problems, such as *density estimation*, *causal inference*, with wide applications in computational biology, such as *single cell genomics*, *multiomics data integration*, and *genomic large language models*.

## Academic Appointment

<b>2025.09-</b>	Assistant Professor - Yale University, New Haven, CT, USA Department of Biostatistics
<b>2021.06-</b> <b>2025.08</b>	Postdoctoral Scholar - Stanford University, Stanford, CA, USA Department of Statistics, advised by Prof. Wing Hung Wong

## Education

<b>2019.09-</b> <b>2021.06</b>	Visiting Ph.D. student - Stanford University, Stanford, CA, USA Department of Statistics, advised by Prof. Wing Hung Wong
<b>2016.09-</b> <b>2019.09</b>	Ph.D. student - Tsinghua University, Beijing, China Department of Automation, advised by Prof. Rui Jiang
<b>2015.08-</b> <b>2016.01</b>	Exchange Student - Lund University, Lund, Sweden Department of Computer Science
<b>2012.09-</b> <b>2016.06</b>	Bachelor in Engineering - Beihang University, Beijing, China ShenYuan Honors College (formerly School of Advanced Engineering)

## Publications

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

## Preprints

- Erpai Luo<sup>†</sup>, **Qiao Liu**<sup>†\*</sup>, Minsheng Hao, Lei Wei, Xuegong Zhang\*. Multi-modal diffusion model with dual-cross-attention for multi-omics data generation and translation[J]. bioRxiv, 2025 (in revision at Nature Communications)
- 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 JASA)
- Qiao Liu**<sup>†</sup>, Wanwen Zeng<sup>†</sup>, 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, 2024.
- 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 papers (as lead author)

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. 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.
8. **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.
9. 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.
10. **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.
11. **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.
12. **Qiao Liu**, Jiaye 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.
13. **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.(conference acceptance rate:20.2%)
14. **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. (conference acceptance rate:18.9%)
15. Pengyu Chen<sup>†</sup>, **Qiao Liu**<sup>†</sup>, Lan Wei, Beier Zhao, Yin Jia, Hairong Lv\*, Xiaolu Fei\*. Automatically structuring on Chinese ultrasound report of cerebrovascular diseases via natural language processing[J]. **IEEE Access**, 2019, 7: 89043-89050.
16. **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.
17. **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.

## Peer-reviewed papers (as co-author)

18. 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.
19. 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.
20. 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.
21. 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.
22. 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.

23. 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.
24. Shuo Li, Weihua Zeng, Xiaohui Ni, **Qiao Liu**,...,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.
25. Shuang Zhang, Yuti Liu, Shuang Chen, **Qiao Liu**, Wanwen Zeng. Applications of Transformer-based language models in bioinformatics: A survey[J]. *Bioinformatics Advances*, 2023.
26. Christopher Lance, Malte D Luecken, Daniel B Burkhardt,...,**Qiao Liu**,...,Fabian J Theis\*. Multimodal single cell data integration challenge: results and lessons learned[J]. Proceedings of Machine Learning Research(*PMLR*), 2022.
27. 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.
28. 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.
29. Qijin Yin, **Qiao Liu**, Zhuoran Fu, Rui Jiang\*. scGraph: a graph neural network-based approach to automatically identify cell types[J]. *Bioinformatics*, 2022.
30. 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.
31. 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. (acceptance rate:26%)
32. 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.
33. 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).
34. 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.
35. 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.
36. 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.
37. 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. (acceptance rate:20.1%)
38. Qijin Yin, Mengmeng Wu, **Qiao Liu**, Rui Jiang\*. DeepHistone: a deep learning approach to predicting histone modifications[J]. *BMC Genomics*, 2019,20(2):193.
39. Shaoming Song, Hongfei Cui, Shengquan Chen, **Qiao Liu**, Rui Jiang\*. EpiFIT: Functional interpretation of transcription factors based on combination of sequence and epigenetic information[J]. *Quantitative Biology*, 2019, 1-11.
40. 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.

## Funding

**NIH 1K99HG013661** (Pathway to Independence Award). Role: PI(sole)

Start/End: Sep 2024 - Sep 2029. Total cost: \$1,068,930

Title: *Bridging the gap between genetic variants and radiomic phenotypes via genomic large language models*

## Invited Talks

2025.08	<i>An AI-powered Bayesian generative modeling approach for causal inference in observational studies</i> JSM (Joint Statistical Meetings), Washington DC, USA
2025.05	<i>Genomic large language models and the applications in Alzheimer's disease</i> STATGEN (Conference on Statistics in Genomics and Genetics), University of Minnesota, USA
2025.03	<i>Multi-modal diffusion model with dual-cross-attention for multi-omics data generation and translation</i> MCBIOS (MidSouth Computational Biology and Bioinformatics Society), University of Utah, USA
2024.11	<i>A Flexible Generative AI Framework for High-dimensional Data Analysis</i> USC Quantitative and Computational Biology Seminar, Los Angeles, USA
2024.10	<i>EpiGePT: a pretrained transformer-based language model for context-specific human epigenomics</i> ICIBM (International Conference on Intelligent Biology and Medicine), Houston, TX, USA
2023.10	<i>A Flexible Generative AI Framework for High-dimensional Data Analysis</i> UNC Biostats Seminar, remote talk
2023.04	<i>CausalEGM: a general causal inference framework by encoding generative modeling</i> UC Berkeley Biostats Seminar, UC Berkeley, USA
2023.04	<i>CausalEGM: a general causal inference framework by encoding generative modeling</i> Tsinghua Statistical Seminar, Tsinghua University (remote)
2022.08	<i>Analyzing spatially resolved transcriptomics data with deep generative model</i> NIH CEGS Annual Meeting, Duke University, USA
2022.08	<i>scTrace: Time series single-cell trajectory inference using deep generative models</i> JSM (Joint Statistical Meetings), Washington DC, USA
2022.06	<i>The application of deep generative models in biology</i> CVI Early Career Research Roundtable, Stanford University, USA
2021.12	<i>Single cell multi-modal integrative analysis with autoencoder</i> NeurIPS conference(The 35th Conference on Neural Information Processing Systems), remote talk
2020.09	<i>DeepCDR: a hybrid graph convolutional network for predicting cancer drug response</i> ECCB conference(The 19th European Conference on Computational Biology), remote talk
2020.03	<i>Simultaneous deep generative modelling and clustering of single-cell genomic data</i> Center for Personal Dynamic Regulomes Seminar (CEGS), Stanford University, USA
2019.07	<i>hicGAN infers super resolution Hi-C data with generative adversarial networks</i> ISMB conference(The 27th Conference on Intelligent Systems for Molecular Biology), Basel, Switzerland
2017.01	<i>A sequence-based method to predict the impact of regulatory variants using random forest</i> APBC conference(The 15th Asia Pacific Bioinformatics Conference), Shenzhen, China

## Honers and Awards

2024.08	<b>Pathway to Independence Award K99/R00</b> , National Human Genome Research Institute (NHGRI)
2021.12	<b>Winner of NeurIPS 2021: Multimodal Single-Cell Data Integration</b> , NeurIPS Competition
2020.08	<b>ECCB Fellowship</b> , International Society for Computational Biology (ISCB)
2019.04	<b>ISMB Travel Fellowship</b> , International Society for Computational Biology (ISCB)
2018.10	<b>National Scholarship</b> , Ministry of Education of China
2015.06	<b>Microsoft Young Fellowship</b> , Microsoft Research Asia

## Teaching Experiences

2025 Spring	Lecturer, <i>STATS 371: Applied Bayesian Statistics</i> Department of Statistics, Stanford University
2024 Spring	Teaching Assistant, <i>STATS 319: Literature of Statistics</i> Department of Statistics, Stanford University
2023 Summer	Teaching Assistant, <i>Introduction to Generative Models</i> Institute for Computational& Mathematical Engineering (ICME), Stanford University
2019 Fall	Teaching Assistant, <i>Intelligent Algorithms and Systems</i> Fundamental Industry Training Center, Tsinghua University
2018 Fall	Teaching Assistant, <i>Statistical Methods with Applications</i> Department of Automation, Tsinghua University
2017 Fall	Teaching Assistant, <i>Introduction to Artificial Intelligence</i> Department of Automation, Tsinghua University

## Professional Activities

<b>Member</b>	<i>American Statistical Association (ASA), International Society of Computational Biology (ISCB)</i> <i>International Chinese Statistical Association (ICSA)</i>
<b>Reviewer</b>	Stats Journals: <i>Journal of the American Statistical Association (JASA), Journal of Causal Inference</i> <i>Journal of Computational and Graphical Statistics, Computational Statistics</i> Comput. Biol. Journals: <i>Nature Methods, Nature Machine Intelligence, Nature Communications,</i> <i>Genome Biology, Bioinformatics, Briefings in Bioinformatics, iScience, Bioinformatics Advances,</i> <i>Genomics, Proteomics &amp; Bioinformatic, BMC Genomics, BMC Bioinformatics, Patterns,</i> <i>Computational Biology, Drug Discovery Today, NAR Genomics and Bioinformatics,</i> <i>Engineering Applications of Artificial Intelligence, Complex &amp; Intelligent Systems,</i> <i>IEEE Transactions on Medical Imaging</i>
<b>PC member</b>	ML conferences <i>ICLR 2025, AAAI 2023-25</i>