# Qiao Liu – Curriculum Vitae

| **Address** | James H. Clark Center | **Telephone** | (650)-283-5489 |
| | Stanford, CA, 94305 | **Email** | liuqiao@stanford.edu |
| **Website** | https://liuq-lab.com | | |

## Research Statement

I'm an incoming 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

**2025.08-** Assistant Professor - Yale University, New Haven, CT, USA
Department of Biostatistics

**2021.06-** Postdoctoral Scholar - Stanford University, Stanford, CA, USA
**2025.07** Department of Statistics, advised by Prof. **Wing Hung Wong**

## Education

**2019.09-** Visiting Ph.D. student - Stanford University, Stanford, CA, USA
**2021.06** Department of Statistics, advised by Prof. Wing Hung Wong

**2016.09-** Ph.D. student - Tsinghua University, Beijing, China
**2019.09** Department of Automation, advised by Prof. Rui Jiang

**2015.08-** Exchange Student - Lund University, Lund, Sweden
**2016.01** Department of Computer Science

**2012.09-** Bachelor in Engineering - Beihang University, Beijing, China
**2016.06** ShenYuan Honors College (Special Pilot Class)

## Publications

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

### Preprints

1. **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.

2. **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, 2024.

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]. medRxiv, 2024.

4. Zhaoyang Zhang, Ziqi Chen, **Qiao Liu**, Jinhan Xie, Hongtu Zhu. Sampling-guided heterogeneous graph neural network with temporal smoothing for scalable longitudinal data imputation. arXiv preprint arXiv:2411.04899, 2024.

5. 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]. bioRxiv, 2024.

## Peer-reviewed papers (as lead author)

6. Zijing Gao†, **Qiao Liu**†\*, Wanwen Zeng, Wing Hung Wong\*, Rui Jiang\*. EpiGePT: a Pretrained Transformer model for human epigenomics[J]. *Genome Biology*, 2024.

7. **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.

8. 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.

9. **Qiao Liu**†, Wanwen Zeng†, 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.

10. Wanwen Zeng†, **Qiao Liu**†, Qijin Yin†, Rui Jiang\*, Wing Hung Wong\*. HiChIPdb: a comprehensive database of HiChIP regulatory interactions[J]. *Nucleic Acids Research*. 2022.

11. **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.

12. **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.

13. **Qiao Liu**, Jiaze Xu, Rui Jiang\*, Wing Hung Wong\*. Density estimation with deep generatative neural networks [J]. Proceedings of the National Academy of Sciences of the United States of America(*PNAS*), 2021, 118(15), e2101344118.

14. **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%)

15. **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 Biology the 18th European Conference on Computational Biology* (*ISMB/ECCB*), 2019. (conference acceptance rate:18.9%)

16. Pengyu Chen†, **Qiao Liu**†, 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.

17. **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.

18. **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)

19. 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.

20. 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.

21. 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.

22. Shuang Zhang, Yuti Liu, Shuang Chen, **Qiao Liu**, Wanwen Zeng. Applications of Transformer-based language models in bioinformatics: A survey[J]. *Bioinformatics Advances*, 2023.

23. 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.

24. 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.

25. 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.

26. Qijin Yin, **Qiao Liu**, Zhuoran Fu, Rui Jiang*. scGraph: a graph neural network-based approach to automatically identify cell types[J]. ***Bioinformatics***, 2022.

27. 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.

28. 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%)

29. 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.

30. 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).

31. 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.

32. 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.

33. 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.

34. 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%)

35. Qijin Yin, Mengmeng Wu, **Qiao Liu**, Rui Jiang*. DeepHistone: a deep learning approach to predicting histone modifications[J]. ***BMC Genomics***, 2019,20(2):193.

36. 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.

37. 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

## Invited Talks

**2024.11**    **USC Quantitative and Computational Biology**, Los Angeles, USA
**2024.10**    **ICIBM 2024**(International Conference on Intelligent Biology and Medicine), Houston, TX, USA
**2023.10**    **UNC Biostats Seminar**, remote talk
**2023.04**    **UC Berkeley Biostats Seminar**, UC Berkeley, USA
**2023.04**    **Tsinghua Statistical Seminar**, Tsinghua University, China (remote)
**2022.08**    **NIH CEGS Annual Meeting**, Duke University, USA

| | |
|---|---|
| **2022.08** | **JSM 2022** (Joint Statistical Meetings), Washington DC, USA |
| **2022.06** | **CVI Early Career Research Roundtable**, Stanford University, USA |
| **2021.12** | **NeurIPS 2021** (The 35th Conference on Neural Information Processing Systems), remote talk |
| **2020.09** | **ECCB 2020** (The 19th European Conference on Computational Biology), remote talk |
| **2020.03** | **Center for Personal Dynamic Regulomes Seminar**, Stanford University, USA |
| **2019.11** | **BIBM 2019** (IEEE International Conference on Bioinformatics & Biomedicine), San Diego, USA |
| **2019.07** | **ISMB 2019** (The 27th Conference on Intelligent Systems for Molecular Biology), Basel, Switzerland |
| **2019.03** | **BUAFAI 2019** (The First Beijing Universities Academic Forum of Artificial Intelligence), Beijing, China |
| **2017.01** | **APBC 2017** (The 15th Asia Pacific Bioinformatics Conference), Shenzhen, China |

## Selected Honers and Awards

| | |
|---|---|
| **2024.08** | **Pathway to Independence Award K99/R00**, NIH NHGRI |
| **2022.12** | **Outstanding Ph.D. Thesis Award**, Beijing Municipal Commission of Education |
| **2020.08** | **ECCB Fellowship**, International Society for Computational Biology (ISCB) |
| **2019.04** | **ISMB Travel Fellowship**, International Society for Computational Biology (ISCB) |
| **2018.10** | **National Scholarship**, Ministry of Education of China (**0.2**% domestically) |
| **2015.06** | **Microsoft Young Fellowship**, Microsoft Research Asia (**40** winners nationwide) |

## Selected Competitions

| | |
|---|---|
| **2021.12** | **NeurIPS 2021 Competition, Multimodal Data Integration**, rank $1^{st}$/2611 in two Joint Embedding Tasks, NeurIPS conference |
| **2021.05** | **RNA Unpaired Probability Prediction Competition**, rank $3^{rd}$/1012, Baidu AI Studio, Baidu Inc |
| **2014.09** | **National** $1^{st}$ **Prize in CUMCM** (Contemporary Undergraduate Mathematical Contest in Modeling), China Society for Industrial and Applied Mathematics |
| **2013&2014** | $1^{st}$ **Prize in NCSMC** (The $5^{th}$&$6^{th}$ National College Students Mathematical Competition), Chinese Mathematical Society |

## Teaching Experiences

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

## Professional Activities

| | |
|---|---|
| **Member** | *American Statistical Association (ASA), International Society of Computational Biology (ISCB)* |
| **Reviewer** | Stats Journals: *Journal of the American Statistical Association (JASA), Journal of Causal Inference Journal of Computational and Graphical Statistics, Computational Statistics* <br> Comput. Biol. Journals: *Nature Methods, Nature Machine Intelligence, Nature Communications, Genome Biology, Bioinformatics, Briefings in Bioinformatics, iScience, Bioinformatics Advances, Genomics, Proteomics & Bioinformatic, BMC Genomics, BMC Bioinformatics, Patterns, Computational Biology, Drug Discovery Today, NAR Genomics and Bioinformatics, Engineering Applications of Artificial Intelligence, Complex & Intelligent Systems, IEEE Transactions on Medical Imaging* |

**PC member** ML conferences *ICLR 2025, AAAI 2023-25*