Xinyi Yu

Ph.D. Candidate, Chinese University of Hong Kong, Shenzhen
Email: 223040233@link.cuhk.edu.cn

EDUCATION

Ph.D., Data Science *2023/09 – Present*

School of Data Science, Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)

Advisor: Prof. Tianwei Yu

M.Phil., Mathematics *2020/09 – 2022/06*

Department of Mathematics, Hong Kong University of Science and Technology (HKUST)

Advisor: Prof. Can Yang

B.Sc., Mathematics *2016/09 – 2020/06*

Cuiying Honors College (Top-Notch Undergraduate Training Program), Lanzhou University

RESEARCH INTERESTS

My research focuses on statistical genetics and computational genomics, with an emphasis on developing scalable statistical and machine learning frameworks for multi-omics integration and genetic analysis of complex diseases. I have worked on cell-type-specific eQTL mapping, spatially resolved multicellular niche detection, and cross-population association analysis, leveraging my interdisciplinary background in statistics, computation, and AI-driven modeling.

PUBLICATIONS

Statistical Genetics and Genomics

[1] **Xinyi Yu**, Xiaomeng Wan, Leqi Tian, Yuheng Chen, Yuyao Liu, Tianwei Yu, Can Yang, Jiashun Xiao (2025). NicheScope: identifying multicellular niches and niche-regulated cell states in spatial transcriptomics. *Submitted*. [preprint][software]

[2] **Xinyi Yu**, Xianghong Hu, Xiaomeng Wan, Zhiyong Zhang, Xiang Wan, Mingxuan Cai, Tianwei Yu, Jiashun Xiao (2025). A unified framework for cell-type-specific eQTLs prioritization by integrating bulk and scRNA-seq data. *American Journal of Human Genetics*. [paper] [software]

[3] **Xinyi Yu**, Jiashun Xiao, Mingxuan Cai, Yuling Jiao, Xiang Wan, Jin Liu, Can Yang (2023). PALM: A powerful and adaptive latent model for prioritizing risk variants with functional annotations. *Bioinformatics*. [paper] [software]

[4] Jiashun Xiao, Mingxuan Cai, **Xinyi Yu**, Xianghong Hu, Gang Chen, Xiang Wan, Can Yang (2022). Leveraging the local genetic structure for trans-ancestry association mapping. *The American Journal of Human Genetic* [paper] [software]

Computational and AI Methods in Biomedial Image Analysis

[1] **Xinyi Yu**, Guanbin Li, Wei Lou, Siqi Liu, Xiang Wan, Yan Chen, Haofeng Li (2023). Diffusion-based data augmentation for nuclei image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [paper] [software]

[2] Wei Lou, **Xinyi Yu**, Chenyu Liu, Xiang Wan, Guanbin Li, Siqi Liu, Haofeng Li (2023). Multi-stream Cell Segmentation with Low-level Cues for Multi-modality Images. *Competitions in Neural Information Processing Systems*. [paper] [software]

Research and Academic Experiences

Algorithm Engineer, Shenzhen Research Institute of Big Data 2022/06 – 2023/06

- Developed a diffusion-based augmentation method for histology image segmentation.
- Published a first-author paper at *MICCAI 2023* and obtained a national patent.
- Achieved 2nd place in the *NeurIPS 2022* multimodal cell segmentation challenge.

Teaching Assistant

STA4100: Statistical Inference II, CUHK-Shenzhen	*2025 Spring*
STA2001: Probability and Statistics II, CUHK-Shenzhen	*2024 Fall*
STA4100: Statistical Inference II, CUHK-Shenzhen	*2024 Spring*
STA2001: Probability and Statistics I, CUHK-Shenzhen	*2023 Fall*
MATH2011: Introduction to Multivariable Calculus, HKUST	*2022 Spring*
MATH1013: Calculus I, HKUST	*2021 Fall*
MATH2011: Introduction to Multivariable Calculus, HKUST	*2021 Spring*

Research Intern

Department of Statistics, Pennsylvania State University *2019/06 – 2020/01*
Advisor: Prof. Qunhua Li

Visiting Student

University of Wisconsin–Madison	*2019/06 – 2020/01*
University of California, Berkeley	*2018/07 – 2018/08*

Honors and Awards

Dean's Fellowship, CUHK-Shenzhen	*2023*
NeurIPS 2022 Medical Image Segmentation Challenge – Winner Finalist Award, SRIBD	*2023*
RedBird PhD Scholarship, HKUST	*2020*
National Scholarship (top 1%), Lanzhou University	*2018*
First-Class Scholarship (top 2%), Lanzhou University	*2017*
Merit Student of Gansu Province, Lanzhou University	*2017*

Technical Skills

Programming & Computing: Python (PyTorch for deep learning), R, Linux
Statistical & Machine Learning: Probabilistic modeling, Bayesian inference, high-dimensional data analysis
Genomics & Bioinformatics: Integrative multi-omics analysis, GWAS, eQTL, spatial transcriptomics