Haotian Cui, Ph.D.

904-832 Bay Street, Toronto ON, Canada M5S 1Z6 | ht.cui@mail.utoronto.ca | (647)904-0252 | Homepage

EDUCATION

07/2024 **Ph.D., Computer Science** **University of Toronto**
Dissertation: Foundation Model for Single-Cell Omics

06/2019 **M.Sc., Biomedical Engineering** **Tsinghua University**

06/2015 **B.Sc., Biomedical Engineering** **Tsinghua University**

EXPERIENCE

07-12/2024 **Worked on an AI-biotech Startup Creation**
Cofounder, led AI-driven drug discovery efforts and fund raising
Full-stack AI for mRNA therapeutics, based on our works of scGPT, AGILE, etc.

04-10/2024 **Part-time research consultant**
Help build up the single-cell foundation model pipeline for vevo.ai
Consulting for scGPT perturbation prediction on the internal assays for Bit.bio

05-12/2021 **Research Intern** **Microsoft Research**
Advisor: Dr. Nan Duan, Dr. Jianfeng Gao
Research on LLM pretraining for coding assistants, optimizing transformer-based models for enhanced code generation and understanding.

2019-2024 **Graduate Student Researcher** **University of Toronto**
Advisor: Prof. Bo Wang **Vector Institute**
Research on large-scale self-supervised models for single-cell omics and drug discovery

RESEARCH INTERESTS

Machine Learning, Genomics, Drug Discovery

- Developing large-scale self-supervised models that enable discovery of biological insights and drug targets from single-cell omics..
- Foundation model and automatic laboratory for therapeutic discoveries, including mRNA drug optimization and delivery vehicle design.
- Multimodal learning for molecular cell biology, for discovery and understanding of generic regulations and multi-faceted developmental or pathological processes.

PUBLICATIONS (700+ citations / Google Scholar / GitHub)

FIRST/CO-FIRST AUTHORSHIP (*equal contribution, first three highlighted, sorted by year afterwards)

1. **Cui, H.***, Wang, C.*, Maan, H., Pang, K., Luo, F., Duan, N. and Wang, B., 2024. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. ***Nature Methods*,** pp.1-11. [link]
   - Pioneering single-cell foundation model with generative pretraining, for building virtual cells.
   - Zero-shot capability of cell type calling and genetic target identification; fine-tuning for perturbation response prediction.
   - Highlighted by Nvidia GTC 2024 [link], NY Times [link], Nature Methods [link, link],

2. **Cui, H.**, Tejada-Lapuerta, A., Brbić, M., Saez-Rodriguez, J., Cristea, S., Goodarzi, H., Lotfollahi, M., Theis, F.J., and Wang, B. 2025. Towards Multimodal Foundation Models in Molecular Cell Biology. ***Nature***. (Forthcoming)
   - Perspective on techniques and data for a multimodal foundation model throughout the central dogma for building virtual cell

3. **Cui, H.***, Xu, Y.*, Pang, K., Li, G., Gong, F., Wang, B., Li, B., 2025. LUMI-lab: a Foundation Model-Driven Autonomous Platform Enabling Discovery of New Ionizable Lipid Designs for mRNA Delivery. ***bioRxiv***, pp.2025-02. (Under review, ***Cell***) [link, video]
   - Fully self-driving lab powered by a new molecular foundation model, conducting experiments in active-learning loops.
   - Discovered new family of ionizable lipids with brominated lipid tails for mRNA delivery
   - Achieved 20.3% gene editing efficiency in lung epithelial cells, the highest reported for inhaled LNP-mediated delivery in mice

4. Wang, C.X.*, **Cui, H.***, Zhang, A.H., Xie, R., Goodarzi, H. and Wang, B., 2025. scGPT-spatial: Continual Pretraining of Single-Cell Foundation Model for Spatial Transcriptomics. ***bioRxiv***, pp.2025-02.

5. Ying, K.*, Song, J.*, **Cui, H.***, Zhang, Y., Li, S., Chen, X., Liu, H., Eames, A., McCartney, D.L., Marioni, R.E. and Poganik, J.R., 2024. MethylGPT: a foundation model for the DNA methylome. ***bioRxiv***, pp.2024-10.

6. Xu, Y.*, Ma, S.*, **Cui, H.***, Chen, J., Xu, S., Gong, F., Golubovic, A., Zhou, M., Wang, K.C., Varley, A. and Lu, R.X.Z., 2024. AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery. ***Nature Communications***, 15(1), p.6305.

7. **Cui, H.***, Maan, H.*, Vladoiu, M.C., Zhang, J., Taylor, M.D. and Wang, B., 2024. DeepVelo: deep learning extends RNA velocity to multi-lineage systems with cell-specific kinetics. ***Genome Biology***, 25(1), p.27.

8. **Cui, H.**, Wang, C., Huang, J., Inala, J.P., Mytkowicz, T., Wang, B., Gao, J. and Duan, N., 2022, December. CodeExp: Explanatory Code Document Generation. In Findings of the Association for Computational Linguistics: ***EMNLP 2022*** (pp. 2342-2354).

9. **Cui, H.***, Wang, C.*, Maan, H., Duan, N. and Wang, B., 2022. scFormer: a universal representation learning approach for single-cell data using transformers. ***bioRxiv***, pp.2022-11.

10. **Cui, H.***, Maan, H.*, Taylor, M.D. and Wang, B., 2022. A Deep Learning Framework for Estimating Cell-specific Kinetic Rates of RNA Velocity. ***The 2022 ICML Workshop on Computational Biology,*** 25(1), p.27.

11. Sugiyama, M.G.*, **Cui, H.***, Redka, D.Y.S., Karimzadeh, M., Rujas, E., Maan, H., Hayat, S., Cheung, K., Misra, R., McPhee, J.B. and Viirre, R.D., 2021. Multiscale interactome analysis coupled with off-target drug predictions reveals drug repurposing candidates for human coronavirus disease. Scientific reports, 11(1), p.23315.

12. Dou, Z.*, **Cui, H.***, Zhang, L. and Wang, B., 2020. Learning global and local consistent representations for unsupervised image retrieval via deep graph diffusion networks. ***arXiv*** preprint arXiv:2001.01284.

CO-AUTHORSHIP

13. Szałata, A., Hrovatin, K., Becker, S., Tejada-Lapuerta, A., **Cui, H.**, Wang, B. and Theis, F.J., 2024. Transformers in single-cell omics: a review and new perspectives. ***Nature Methods***, 21(8), pp.1430-1443.

14. Tan, R., **Cui, H.**, Wang, B., Perez, K., Costa, A.D., Jordan, A., Karacic, T., Elganainy, D., Gui, D.Y., Kim, S. and Yuan, C., 2024. Abstract A029: Evaluating and interpreting scGPT: A foundation model for single-cell biology in real-world cancer clinical trial data. ***Cancer Research***, 84(17_Supplement_2), pp.A029-A029.

15. Hu, H., Ma, Y., Gao, X., Song, D., Li, M., Huang, H., Qian, X., Wu, R., Shi, K., Ding, H., Lin, M., Chen, X., Zhao, W., Qi, B., Zhou, S., Chen, R., Gu, Y., Chen, Y., Lei, Y., Wang, C., Wang, C., Tong, Y., **Cui, H.**, Abdal, A., Zhu, Y., Tian, X., Chen, Z., Lu, C., Yang, X., Mu, J., Lou, Z., Eghtedari, M., Zhou, Q., Oberai, A. and Xu, S. 2023. Stretchable ultrasonic arrays for the three-dimensional mapping of the modulus of deep tissue. ***Nature Biomedical Engineering***, 7(10), pp.1321-1334.

16. Zaidi, M., **Cui, H.**, Wang, B., McKee, T.D. and Wouters, B.G., 2021. Abstract PO-018: Computational staining of tumor hypoxia from H&E images using convolutional neural networks. ***Clinical Cancer Research***, 27(5_Supplement), pp.PO-018.

17. Huang, J., Wang, C., Zhang, J., Yan, C., **Cui, H.**, Inala, J.P., Clement, C., Duan, N. and Gao, J., 2022. Execution-based Evaluation for Data Science Code Generation Models. ***DaSH 2022***, p.28.

18. Yang, J., **Cui, H.**, Li, S., Gao, S., Guo, J. and Lu, Z., 2020. Outline Extraction with Question-Specific Memory Cells. ACM *Transactions on Asian and Low-Resource Language Information Processing (**TALLIP**),* 19(4), pp.1-17.

19. Liu, X., Mou, L., **Cui, H.**, Lu, Z. and Song, S., 2020. Finding decision jumps in text classification. ***Neurocomputing***, 371, pp.177-187.

20. Liu, X., Mou, L., **Cui, H.**, Lu, Z. and Song, S., 2018. Jumper: Learning when to make classification decisions in reading. ***arXiv*** preprint arXiv:1807.02314.

21. Li, G., Deng, L., Tian, L., **Cui, H.**, Han, W., Pei, J. and Shi, L., 2018. Training deep neural networks with discrete state transition. ***Neurocomputing***, 272, pp.154-162.

22. Lu, Z., Liu, X., ***Cui, H.***, Yan, Y. and Zheng, D., 2018, July. Object-oriented Neural Programming (OONP) for Document Understanding. In Proceedings of the 56th ***Annual Meeting of the Association for Computational Linguistics*** (Volume 1: Long Papers) (pp. 2717-2726).

PATENTS

- Self-Driving System and Method for AI-Based Design and Discovery of Lipid Nanoparticles as Therapy Delivery Vehicles (Provisional Application, USPTO 63/729,809)
- Systems and Methods for AI-Based Ionizable Lipid Development for Lipid Nanoparticle (LNP)-Based Cargo Molecule Delivery (PCT/CA2024/050644)

PROFESSIONAL SERVICE

2019-now	Peer reviewer of journals and conferences, including **ACL, EMNLP, ICLR, IEEE Transactions on Signal and Information Processing over Networks**
2022	**Program committee member for ACL2022 DialDoc Workshop [**link**]**

PRESENTATIONS AND INVITED TALKS

2023	"scGPT: towards building a foundation model for single-cell multi-omics using generative AI", School of Medicine Single-Cell Club, New York University
2023	scGPT webinar, Superbio.ai
2023	"scGPT: towards building a foundation model for single-cell multi-omics using generative AI", CEGS_ICA Journal Club, New York Genome Center
2022	"CodeExp: Explanatory Code Document Generation", EMNLP 2022
2022	"A Deep Learning Framework for Estimating Cell-specific Kinetic Rates of RNA Velocity", The 2022 ICML Workshop on Computational Biology

AWARDS AND HONORS

2024	**Department of Computer Science Graduate Program Award**, University of Toronto
2022	**Best Poster Award**, ICML 2022 Workshop on Computational Biology
2014	**First Prize, Texas Instrument National Biomedical Electronics Competition**
2014	**Science and Technology Scholarship**, Tsinghua University
2013	**First Prize, NI National Virtual Instruments Competition**