

Ziheng Duan

Department of Computer Science
Donald Bren School of Information and Computer Sciences
University of California, Irvine

duanziheng1206@gmail.com
+1 424 239 4085
zihengduan.cn

EDUCATION

- Ph.D. Computer Science, University of California, Irvine, 2021 - 2025
M.S. Computer Science, University of California, Irvine, 2021 - 2024
B.S. Automation & Computer Science (Dual Degree), Zhejiang University, 2016 - 2020

RESEARCH INTERESTS

AI Methodologies: Graph Learning, Multimodal Modeling, Large Language Models
Biological Research: Single-Cell Genomics, Spatial Transcriptomics, Drug Discovery

WORK EXPERIENCE

Postdoctoral Researcher Mar. 2026 – Oct. 2026 (expected)
University of California, Irvine, CA

- Developed TargetSage, an LLM reasoning agent for genome-scale therapeutic target identification that combines agentic evidence retrieval, GRPO-trained gene-level reasoning, and positive-unlabeled learning to correct false negatives among unlabeled druggable genes; outperforms ten baselines across 15 benchmark tasks covering 19,032 human protein-coding genes and prospectively ranks targets that later reached the clinic.
- Developed CellReprogrammer, a differentiable framework that optimizes sparse gene-perturbation sets to shift disease-associated cell states toward health, coupling a disease-state classifier with a foundation-model perturbation surrogate and relaxed top- K selection while co-optimizing druggability and on-target safety; evaluated across five psychiatric disorders and three brain cell types.

Machine Learning Research Intern Jun. 2025 – Present
Genesis Therapeutics, San Diego, CA

- Developed KNOT: A Knowledge Graph and Omics Integration Framework for Target Identification, integrating multi-omics data and biomedical knowledge graphs to enhance drug target prediction.
- Built an autoimmune single-cell reference atlas spanning blood, gut, and skin across four studies (over 8M cells, 1,200+ donors) to profile target and pathway expression by immune cell type and disease, pairing donor-level statistical validation with eQTL–GWAS colocalization to support target evaluation.

Graduate Researcher Sep. 2021 – Jun. 2025
University of California, Irvine, CA
Developed trustworthy computational methods to understand single-cell gene regulation, especially in spatial contexts, and their contributions to various diseases.

PUBLICATIONS

First Author Publications

- [10] **Duan, Z.**, Li, X., Xiao, Z., Ying, R., & Zhang, J. (2025). MUSE: A Multi-slice Joint Analysis Method for Spatial Transcriptomics Experiments. *CIKM-2025: The 34th ACM International Conference on Information and Knowledge Management*. doi:10.1145/3746252.3761240
- [9] **Duan, Z.**, Li, X., Zhang, Z., Song, J., & Zhang, J. (2025). DISCO: A Diffusion Model for Spatial Transcriptomics Data Completion. *ICIP-2025: 2025 IEEE International Conference on Image Processing*. doi:10.1109/ICIP55913.2025.11084277
- [8] **Duan, Z.**, Xu, S., Lee, C.Y., Riffle, D., & Zhang, J. (2024). iMIRACLE: an Iterative Multi-View Graph Neural Network to Model Intercellular Gene Regulation from Spatial Transcriptomic Data. *CIKM-2024: The 33rd ACM International Conference on Information and Knowledge Management*. doi:10.1145/3627673.3679574
- [7] **Duan, Z.**, Riffle, D., Li, R., Liu, J., Min, M. R., & Zhang, J. (2024). Impeller: a path-based heterogeneous graph learning method for spatial transcriptomic data imputation. *Bioinformatics*, btae339. doi:10.1093/bioinformatics/btae339
- [6] **Duan, Z.**, Xu, S., Sai Srinivasan, S., Hwang, A., Lee, C. Y., Yue, F., ... & Zhang, J. (2024). scENCORE: leveraging single-cell epigenetic data to predict chromatin conformation using graph embedding. *Briefings in Bioinformatics*, 25(2), bbae096. doi:10.1093/bib/bbae096
- [5] **Duan, Z.**, Lee, C., & Zhang, J. (2023). ExAD-GNN: explainable graph neural network for Alzheimer's disease state prediction from single-cell data. *APSIPA Transactions on Signal and Information Processing*, 12(5). doi:10.1561/116.00000239
- [4] **Duan, Z.**, Dai, Y., Hwang, A., Lee, C., Xie, K., Xiao, C., ... & Zhang, J. (2023). iHerd: an integrative hierarchical graph representation learning framework to quantify network changes and prioritize risk genes in disease. *PLoS Computational Biology*, 19(9), e1011444. doi:10.1371/journal.pcbi.1011444
- [3] **Duan, Z.**, Xu, H., Wang, Y., Huang, Y., Ren, A., Xu, Z., ... & Wang, W. (2022). Multivariate time-series classification with hierarchical variational graph pooling. *Neural Networks*, 154, 481-490. doi:10.1016/j.neunet.2022.07.032
- [2] **Duan, Z.**, Xu, H., Huang, Y., Feng, J., & Wang, Y. (2022). Multivariate time series forecasting with transfer entropy graph. *Tsinghua Science and Technology*, 28(1), 141-149. doi:10.26599/TST.2021.9010081
- [1] **Duan, Z.**, Wang, Y., Ye, W., Fan, Q., & Li, X. (2022). Connecting latent relationships over heterogeneous attributed network for recommendation. *Applied Intelligence*, 52(14), 16214-16232. doi:10.1007/s10489-022-03340-7

Other Publications

- [23] Lee, C. Y., Hwang, A., ..., **Duan, Z.**, ..., & Girgenti, M. J. (2026). Central amygdala single-nucleus atlas reveals chromatin and gene transcription dynamics in human alcohol use disorder. *Nature Communications*. doi:10.1038/s41467-026-68351-1
- [22] Zou, Z. & **Duan, Z.** (2025). SpaLLM: a general framework for spatial domain identification with large language models. *Frontiers in Bioinformatics*. 5, 1713975. doi:10.3389/fbinf.2025.1713975

- [21] Gaynor-Gillett, S C, ..., **PsychENCODE Consortium**, ... & White, K. (2025). A map of enhancer regions in primary human neural progenitor cells using capture STARR-seq. *Genome Research*. 35(8), 1887-1901. doi:10.1101/gr.279584.124
- [20] Hwang, A., Skarica, M., ..., **Duan, Z.**, ..., Zhang, J., & Girgenti, M. J. (2025). Single-cell transcriptomic and chromatin dynamics of the human brain in PTSD. *Nature*. doi:10.1038/s41586-025-09083-y
- [19] Liu, J., Xu, S., Riffle, D., **Duan, Z.**, Min, M. R., & Zhang, J. (2024). Understanding Transcriptional Regulatory Redundancy by Learnable Global Subset Perturbations. *ACML-2024: Proceedings of the 16th Asian Conference on Machine Learning*, PMLR 260:383–398. openreview.net/forum?id=2ODIoB7Ijg
- [18] Li, X., **Duan, Z.**, Dai, Y., Xu, S., & Zhang, J. (2024). iHAST: Integrating hybrid attention for super-resolution in spatial transcriptomics. In *Proceedings of the 35th British Machine Vision Conference (BMVC 2024)*. [PDF]
- [17] Dai, Y., **Duan, Z.**, Xu, S., & Zhang, J. (2024). CytoFlow: a novel computational method to construct signal transduction networks at single-cell resolution based on flow networks. *MedAI-2024: IEEE International Conference on Medical Artificial Intelligence*, 212–217. doi:10.1109/MedAI62885.2024.00034
- [16] Wen, C., Margolis, M., Dai, R., Zhang, P., Przytycki, P. F., Vo, D. D., ..., **PsychENCODE Consortium**, ... & Gandal, M. J. (2024). Cross-ancestry atlas of gene, isoform, and splicing regulation in the developing human brain. *Science*, 384(6698), eadho829. doi:10.1126/science.adho829
- [15] Deng, C., Whalen, S., Steyert, M., Ziffra, R., Przytycki, P. F., Inoue, F., ..., **PsychENCODE Consortium**, ... & Pollard, K. S. (2024). Massively parallel characterization of regulatory elements in the developing human cortex. *Science*, 384(6698), eadh0559. doi:10.1126/science.adh0559
- [14] Huuki-Myers, L. A., Spangler, A., Eagles, N. J., Montgomery, K. D., Kwon, S. H., Guo, B., ..., **PsychENCODE Consortium**, ... & Maynard, K. R. (2024). A data-driven single-cell and spatial transcriptomic map of the human prefrontal cortex. *Science*, 384(6698), eadh1938. doi:10.1126/science.adh1938
- [13] Ruzicka, W. B., Mohammadi, S., Fullard, J. F., Davila-Velderrain, J., Subburaju, S., Tso, D. R., ..., **PsychENCODE Consortium**, ... & Kellis, M. (2024). Single-cell multi-cohort dissection of the schizophrenia transcriptome. *Science*, 384(6698), eadg5136. doi:10.1126/science.adg5136
- [12] Emani, P. S., Liu, J. J., Clarke, D., Jensen, M., Warrell, J., Gupta, C., ..., **Duan, Z.**, ... & Gerstein, M. (2024). Single-cell genomics and regulatory networks for 388 human brains. *Science*, 384(6698), eadi5199. doi:10.1126/science.adi5199
- [11] Dai, R., Chu, T., Zhang, M., Wang, X., Jourdon, A., Wu, F., ..., **PsychENCODE Consortium**, ... & Liu, C. (2024). Evaluating performance and applications of sample-wise cell deconvolution methods on human brain transcriptomic data. *Science Advances*, 10(21), eadh2588. doi:10.1126/sciadv.adh2588
- [10] Pratt, H. E., Andrews, G., Shedd, N., Phalke, N., Li, T., Pampari, A., ..., **PsychENCODE Consortium**, ... & Weng, Z. (2024). Using a comprehensive atlas and predictive models to reveal the complexity and evolution of brain-active regulatory elements. *Science Advances*, 10(21), eadj4452. doi:10.1126/sciadv.adj4452
- [9] Xia, Y., Xia, C., Jiang, Y., Chen, Y., Zhou, J., Dai, R., ..., **PsychENCODE Consortium**, ... & Chen, C. (2024). Transcriptomic sex differences in postmortem brain samples from patients with psychiatric disorders. *Science Translational Medicine*, 16(749), eadh9974. doi:10.1126/scitranslmed.adh9974

- [8] Lee, C. Y., Riffle, D., Xiong, Y., Momtaz, N., Lei, Y., Pariser, J. M., ..., **Duan, Z.**, ... & Zhang, J. (2024). Characterizing dysregulations via cell-cell communications in Alzheimer's brains using single-cell transcriptomes. *BMC neuroscience*, 25(1), 24. doi:10.1186/s12868-024-00867-y
- [7] Zhang, R., **Duan, Z.**, Lee, C. Y., Riffle, D., Min, M. R., & Zhang, J. (2023). Turtling: A Time-aware Neural Topic Model on NIH Grant Data. *Bioinformatics Advances*, vbado96. doi:10.1093/bioadv/vbado96
- [6] Lee, C. Y., Chen, Y., **Duan, Z.**, Xu, M., Girgenti, M. J., Xu, K., ... & Zhang, J. (2022). Venus: An efficient virus infection detection and fusion site discovery method using single-cell and bulk RNA-seq data. *PLOS Computational Biology*, 18(10), e1010636. doi:10.1371/journal.pcbi.1010636
- [5] Wang, Y., **Duan, Z.**, Huang, Y., Xu, H., Feng, J., & Ren, A. (2022). MTHetGNN: A heterogeneous graph embedding framework for multivariate time series forecasting. *Pattern Recognition Letters*, 153, 151-158. doi:10.1016/j.patrec.2021.12.008
- [4] Xu, H., **Duan, Z.**, Wang, Y., Feng, J., Chen, R., Zhang, Q., & Xu, Z. (2021). Graph partitioning and graph neural network based hierarchical graph matching for graph similarity computation. *Neurocomputing*, 439, 348-362. doi:10.1016/j.neucom.2021.01.068
- [3] Liao, Z., Yang, Y., Sun, C., Wu, R., **Duan, Z.**, Wang, Y., ... & Xu, J. (2021). Image-based prediction of granular flow behaviors in a wedge-shaped hopper by combining DEM and deep learning methods. *Powder Technology*, 383, 159-166. doi:10.1016/j.powtec.2021.01.041
- [2] Wang, B., Zhou, C., **Duan, Z.**, Zhu, Q., Wu, J., & Xiong, R. (2020). Untethered quadrupedal hopping and bounding on a trampoline. *Frontiers of Mechanical Engineering*, 15, 181-192. doi:10.1007/s11465-019-0559-5
- [1] Wang, Y., **Duan, Z.**, Liao, B., Wu, F., & Zhuang, Y. (2019, July). Heterogeneous attributed network embedding with graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 10061-10062). doi:10.1609/aaai.v33i01.330110061

INVITED TALKS

- “scENCORE, a computationally method to reconstruct chromatin compartments from the more affordable and widely accessible single-cell epigenetic data.”
13th Annual Symposium, Southern California Systems Biology May 2024
- “iHerd, a hierarchical graph representation learning method for identifying key driver genes by analyzing gene network changes across different conditions.”
2023 SCORCH Consortium Meeting Dec 2023

AWARDS AND HONORS

- Travel Grant, IEEE International Conference on Image Processing (ICIP 2025) 2025
- Dean's Fellowship, University of California, Irvine 2021
- Research and Innovation Scholarship, Zhejiang University 2019
- Academic Excellence Scholarship, Zhejiang University 2017 – 2019

TEACHING

- *Reader*, University of California, Irvine Spring 2022

ICS 6D: Discrete Mathematics for Computer Science

- *Reader*, University of California, Irvine
ICS 178: Machine Learning and Data Mining
- *Reader*, University of California, Irvine
ICS 178: Machine Learning and Data Mining

Winter 2021

Fall 2021

MENTORING

- *Mentor*, Google Summer of Code 2026 Summer 2026
StaR: A Stability-Aware Representation Learning Framework for Spatial Domain Identification
Guided the development of StaR, a plug-in training framework for improving the stability and reproducibility of spatial domain identification across random initializations.
- *Mentor*, Google Summer of Code 2025 Summer 2025
RAG-ST: Retrieval-Augmented Generation for Spatial Transcriptomics
Guided the development of a retrieval-augmented framework for predicting spatial gene expression from histology images, focusing on interpretability and modular design.
- *Mentor*, Google Summer of Code 2024 Summer 2024
BenchmarkST: Spatial Transcriptomics Gene Imputation Benchmarking
Supervised the creation of a benchmarking suite for evaluating cross-platform, multi-species gene expression imputation methods.

SERVICES

Journal Reviewer

Applied Soft Computing
Big Data Mining and Analytics
Big Data Research
Biochemical Genetics
BioData Mining
BMC Bioinformatics
BMC Biology
BMC Genomics
BMC Medical Imaging
BMC Medicine
Computers in Biology and Medicine
Computing
Data Mining and Knowledge Discovery
Digital Signal Processing
Discover Artificial Intelligence
Discover Computing
Engineering Applications of Artificial Intelligence
Expert Systems With Applications

Functional & Integrative Genomics
Gene
Genome Medicine
Heliyon
IEEE Access
IEEE Transactions on Knowledge and Data Engineering
IEEE Transactions on Pattern Analysis and Machine Intelligence
Information Science
International Journal of Machine Learning and Cybernetics
Journal of Experimental & Theoretical Artificial Intelligence
Journal of Hydrology
Knowledge and Information Systems
Knowledge Based Systems
Nature Communications
Neural Networks
Neurocomputing
npj Systems Biology and Applications
Pattern Recognition
Pattern Recognition Letters
Reliability Engineering & System Safety
Scientific Reports
The Journal of Supercomputing

Conference Reviewer

IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
IEEE International Conference on AI for Medicine, Health, and Care (AIMHC)
Learning on Graphs Conference (LoG)
The International Conference on Intelligent Systems for Molecular Biology (ISMB)

Updated September 2026