Mehrdad Zandigohar

Chicago, IL 60612 | mzandi2@uic.edu | (773) 677-9079 | LinkedIn | GitHub | Google Scholar

PhD candidate in Bioinformatics (Graduation in Dec. 2025) with expertise in generative models (GenAI), transformers, variational autoencoders (VAEs), and retrieval-augmented generation (RAGs). Applied domains are gene regulatory networks (GRNs), transcription factor inference, and single-cell and spatial genomics.

Education

University of Illinois Chicago, Chicago, IL   8/2020 – present
Ph.D., Bioinformatics, Biomedical Engineering Department (GPA 3.75/4.00)

Southern Illinois University Edwardsville, Edwardsville, IL   8/2018 – 8/2020
M.Sc., Mechanical Engineering (GPA 3.8/4.0)

Sharif University of Technology, Tehran, Iran   9/2012 – 7/2017
B.Sc., Chemical Engineering (Major), B.Sc., Mechanical Engineering (Minor)

Skills

- **Data Science & Machine Learning:** Deep learning, generative models, retrieval-augmented generation (RAGs), statistical modeling, data mining
- **Programming:** Python, R, Bash, MATLAB, C/C++
- **Technologies:** PyTorch (CUDA), Seurat, Scanpy, BioBERT, LlamaIndex, Hugging Face
- **Bioinformatics:** Genomics, gene regulatory networks (GRNs), scRNA-seq, scATAC-seq, sc-multi-omics, spatial transcriptomics, decontamination, cell-cell communication, pseudo-time trajectory inference, GO enrichment analysis, GSEA
- **Tools & Infrastructure:** Git, tmux, Conda environments, Jupyter Notebooks, AWS, Docker
- **Data Visualization:** matplotlib, ggplot2, seaborn

Research & Development

scRegulate: Interpretable Generative Model for Transcription Factor Activity
Graduate Research Assistant – University of Illinois Chicago   *2023 – 2025*

- Developed **scRegulate**, **a biologically grounded variational autoencoder** (VAE) for inferring transcription factor (TF) activity from single-cell RNA-seq data, embedding gene regulatory networks (TF–target and TF–TF priors) into an interpretable latent space where each neuron corresponds to a known TF.
- Integrated **ULM-based initialization**, **sparse regularization**, and a **dynamic prior decay** mechanism to balance data-driven inference with biological priors.
- Outperformed leading tools by 10–50 % across multiple datasets (experimental, perturb-seq, synthetic) using metrics such as **AUROC (0.71–0.86)**, **AUPRC (0.80–0.95)**, **ARI**, **NMI** and **F1**, and accurately recapitulated known perturbation effects (e.g., ELK1, EGR1). Remained robust under 10-30% dropout noise.
- Released as a fully open-source Python package with extensive engineering:
    - Designed the tool's **logo and visual identity**
    - Built **automated CI/CD pipelines** using **GitHub Actions**
    - Packaged for **PyPI** and **Conda**
    - Created **interactive tutorials** and full documentation via **GitHub Pages**
    - Added licensing, Code of Conduct, and issue templates for user support

First-author bioRxiv preprint (2025). Under review (*Bioinformatics)*. Available at: github.com/YDaiLab/scRegulate

RAGulate: Retrieval-Augmented Generation for Literature-Guided Genomics
PhD Candidate – University of Illinois Chicago   *2025 – Present*

- Designing **RAGulate** (in Python), a retrieval-augmented framework that integrates LLMs (e.g., BioGPT, PubMedBERT) and GPU-accelerated semantic search (e.g., NeMo Retriever, cuVS) to validate and prioritize transcription factor-target gene predictions with literature-grounded evidence.
- Built a literature mining pipeline using PubMed APIs, TF and gene name disambiguation, semantic reranking, and abstract-level relevance scoring to retrieve context-specific evidence.
- Integrating LLM-driven retrieval with gene regulatory priors (e.g., TFLink, GTRD) to enhance inference of TF-target relationships in spatial, disease-specific, or under-annotated single-cell datasets.
- Implementing as a modular Python toolkit, with plans for GitHub-based release, documentation, and tutorials following the scRegulate packaging model.
- Currently under development, with benchmarking and manuscript preparation in progress.

Information Retrieval in Sparse & High-Dimensional Genomic Data

Graduate Research Assistant – University of Illinois Chicago   *2021 – 2023*

- Investigated information retrieval challenges (in Python and R) in scATAC-seq, a sparse and high-dimensional modality, by applying and comparing various TF-IDF transformation strategies tailored to genomics data.
- Evaluated the impact of different latent inference models, including PCA, LSI, autoencoders (AE), variational autoencoders (VAE), sparse VAE, and stacked VAE, on the quality of extracted biological signals.
- Systematically benchmarked cases using unsupervised clustering metrics (ARI, NMI, F1), classification accuracy, and empirical significance testing. Published as first-author in the IEEE BIBM 2022.

Leadership & Technical Operations

Leadership in Lab Computing & Mentorship

University of Illinois Chicago   *2022 – Present*

- **Infrastructure & Operations:** Configured remote access; maintained 3 multi-user GPU servers (CUDA/Conda, upgrades) supporting ≈10 active researchers; managed job scheduling and user access.
- **Procurement & Grants:** Oversaw hardware procurement and setup; wrote and secured an NVIDIA GPU grant to expand deep-learning capabilities.
- **Mentorship & Collaboration:** Mentored 10+ grads/undergrads on GPU use, coding best practices, and project workflows; developed internal documentation and streamlined reproducible workflows.

Leadership & Service: Graduate Student Council Member

University of Illinois Chicago   *2022 – 2023*

- Contributed to policy discussions, funding decisions, and advocated graduate students' needs across departments.

Notable Projects

Single-Cell Atlas of Uterine Fibroids and Myometrium (UCSF Collab)   *2021 – 2022*

- Analyzed 96,573 single-cell RNA-seq profiles from fibroid and myometrium tissue in a UCSF collaboration.
- Mapped 16 cell types, reconstructed lineages, and modeled cell-cell communication, revealing new fibroblast and lymphatic-endothelial subtypes linked to fibroid growth.
- Co-authored resulting publication: *Human Reproduction*, cited 41+ times

Multi-omics and Xenium Analysis in Muscle Hernia Model (Northwestern Collab)   *2023 – 2025*

- Processed and analyzed multi-omics/spatial datasets (scRNA-seq, scATAC-seq, Xenium) in mouse models of inguinal hernia. Pinpointed a Pgr^+ fibroblast subtype with TGFb signaling that drives fibrosis and atrophy.
- Handled GEO data deposition, metadata curation, and reproducibility. Co-authored *JCI Insight*.

Spatial Transcriptomics in MED12/HMGA2 Fibroids (Northwestern Collab)   *2022 – 2024*

- Integrated spatial transcriptomics with matched scRNA-seq data from uterine fibroid samples.
- Conducted cell-type deconvolution, spatially resolved ligand-receptor analysis, and subtype comparisons across genetic variants (MED12 vs. HMGA2).
- Revealed spatial signaling rewiring and tumor-specific pathways; results contributed to manuscript.

Benchmarking ML Classifiers for Breast-Cancer Diagnosis   *2020 – 2021*

- Benchmarked ANN, SVM, RF, AdaBoost, GP & QDA on 569 breast-cancer FNA samples (QDA AUC 0.97)
- Feature analysis pinpointed fractal dimension and compactness as the strongest malignancy indicators.

Select Publications

- **Zandigohar M**, Rehman J, Dai Y. *scRegulate: Single-Cell Regulatory-Embedded Variational Inference of Transcription Factor Activity from Gene Expression.* bioRxiv, 2025.
- **Zandigohar M**, Pang J, Rodrigues A, Roberts RE, Dai Y, Koh TJ. *Transcription Factor Activity Regulating Macrophage Heterogeneity during Skin Wound Healing. J Immunol.* 2024;213(4):506–518.
- **Zandigohar M**, Dai Y. *Information retrieval in single cell chromatin analysis using TF-IDF transformation methods. IEEE BIBM*, 2022: 877–882.
- Goad J, Rudolph J, **Zandigohar M**, Tae M, Dai Y, Wei JJ, Bulun SE. *Single-cell sequencing reveals novel cellular heterogeneity in uterine leiomyomas. Hum Reprod.* 2022;37(10):2334–2349.

Honors and Awards

- Top Read Article, The Journal of Immunology: "Transcription Factor Activity Regulating Macrophage Heterogeneity during Skin Wound Healing"   2024
- 3rd Place, Oral Presentation – 5th Annual BME Research Symposium, UIC   2023
- Graduate Student Council (GSC) Award – University of Illinois at Chicago   2023
- NSF Student Travel Awards – ACM-BCB & IEEE BIBM [AWD-003814; AWD-2131662]   2022
- 2nd Place, Oral Presentation – 2nd Annual CBQB Research Day, UIC   2022