

# Wu, Renjie (Roger)

Email: [renjiwu@hsph.harvard.edu](mailto:renjiwu@hsph.harvard.edu) | Phone: +1 (857) 209-1097  
LinkedIn: [linkedin.com/in/renjie-wu-334199325/](https://www.linkedin.com/in/renjie-wu-334199325/) | GitHub: [github.com/vo-olb/](https://github.com/vo-olb/)

## EDUCATION

### Harvard University, School of Public Health

Boston, MA

M.S. in Computational Biology

08/2024 – 05/2026 (Expected)

· **In-Progress Coursework:** Architecting & Engineering Software Systems, Applied Regression Analysis, Data Science

### Peking University, School of Life Sciences

Beijing, CN

B.Eng. in Bioinformatics

09/2020 – 07/2024

· **Major GPA:** 3.8/4.0, **Ranking:** 1<sup>st</sup> in the major

· **Major Coursework:** Data Structures and Algorithms (98.5), Probability and Mathematical Statistics (98), Statistical Analysis of Genomics Data (98.3), Bioinformatics Lab (A), Genetics (99)

## SKILLS

- **Fields:** Data Science (data scraping and handling, statistical modeling, deep learning), Statistical Analysis (hypothesis testing, Bayesian inference), Computational Genomics (RNA-seq data analysis), Software Development (coding, design principles), Software Engineering (dev cycles concept)
- **Programming:** [Proficient] Python, R, Bash, [Experienced] C/C++, SQL, HTML, JavaScript
- **Tools:** [Proficient] VS Code, [Experienced] HPC, Git, Docker
- **Packages:** [Data Science] Pandas, Sklearn, Pytorch, Matplotlib, [Genomics Analysis] DESeq, GSEA, Seurat, Scanpy
- **Language:** [Native] Mandarin, [Proficient] English

## WORK EXPERIENCES

### Inferring dynamic gene regulatory networks (dGRN) based on single-cell trajectories

Beijing, CN

Research Assistant in Dr. Wang Jinzhao's Lab, Peking University

11/2023 – 05/2024

- Collected datasets and developed a data processing pipeline, a deep learning model, and benchmarks to predict dGRN, based on literature review
- Designed, implemented and optimized the deep learning model (two-stage attention layer, GNN and MLP) in *Pytorch* framework on single-cell RNA-seq data of mouse embryonic stem cells to predict their dGRN, achieving an AUC of 0.78 and surpassing prior records ( $<0.53$ )
- Evaluated and visualized the model prediction efficacy through both numerical indicators (loss function, AUC) and biological validations (statistical enrichment analysis by *ToppGene*, trajectory analysis by *Scvelo*)
- Analyzed the effects of different components in the model architecture by ablation experiments

### Decoding gene-gene interactions by predicting target gene expression from regulatory genes

Remote

Undergraduate Researcher in Dr. Cao Junyue's Lab, The Rockefeller University

06/2023 – 10/2023

- Designed, implemented and optimized an auto-encoder model in *TensorFlow* framework on RNA-seq data of ~1.5 million mature mouse brain cells to predict target gene expression
- Experimented on various combinations of data preprocessing methods (cell aggregation, library size normalization, logarithmic transformation and scaling) using *Scanpy*, and achieved better prediction versus standard pipeline
- Benchmarked the auto-encoder model versus other models (naïve model, linear model and ridge model) using *Sklearn* in their predictive accuracy, robustness, and computational efficiency

### Investigating Epithelial-Mesenchymal Plasticity (EMP) in colorectal cancer (CRC)

Beijing, CN

Research Assistant in Dr. Deng Hongkui's and Dr. Cheng Li's Lab, Peking University

11/2022 – 05/2023

- Collected various types of genomic features related to EMP (a biological phenomenon) in CRC by literature review
- Developed and executed RNA-seq analysis pipeline (including data preprocessing, batch effect handling, feature engineering, clustering, visualization) on clinical data cohorts of ~50,000 CRC epithelial cells, and annotated cells statistically associated with these features using *Seurat*
- Conducted RNA-seq analyses to showcase the validity of our lab-built CRC model to recapitulate the EMP identified from clinical data, using both bulk data analyses (cell type deconvolution by *MuSiC*, statistical enrichment analysis by *GSEA*) and single-cell data analyses (mNN label transfer, differential expression analysis, AUC analysis by *Seurat*)
- Analyzed RNA-seq data of ~10,000 cells of the lab-built model, inferred the origins of the EMP by *Scvelo*, and identified specific regulatory genes contributing to the EMP by *SCENIC*

## ACHIEVEMENTS

- **Publication:** "Multiomics dynamic learning enables personalized diagnosis and prognosis for pancancer and cancer subtypes." *Briefings in bioinformatics*, 2023.
- **Awards:** Harvard Chan Scholarship (2024), Peking University Award for Contribution in Student Organizations (2022), Peking University Qin Wanshun -- Jin Yunhui Scholarship (2021)