Shijie Tang

(412) 980-3785 | shijiet@andrew.cmu.edu | Linkedin | GitHub

EDUCATION

Carnegie Mellon University, School of Computer Science Pittsburgh, PA
Master of Science in Computational Biology May 2026
Courses: Generative AI, Algorithm and Advanced Data structure, String Algorithm, Machine Learning for Scientists

Zhejiang University, Zhejiang University - University of Edinburgh (ZJE) Institute Haining, China
Dual Bachelor Degree of Science in Bioinformatics with Honours June 2024
Courses: Database and Software Technology, Biomedical Informatics, Computational Biology and System Biology
Awards: Excellent League Cadre, Excellent League Member and Student Leadership Award

PUBLICATION

- Y Tang, **S Tang**, W Yang et al. "MED12 loss activates endogenous retroelements to sensitise immunotherapy in pancreatic cancer" Gut (2024) DOI: 10.1136/gutjnl-2024-332350
- J Li, L Liang, S Du, **S Tang**, H Lai, C Kingsford. "ARCADE: Controllable Codon Design from Foundation Models via Activation Engineering" Pre-print (2025) DOI: 10.1101/2025.08.19.668819

Research Experience and Internships

Carnegie Mellon University Pittsburgh, PA
Research Assistant, *Supervised by Prof. Carl Kingsford* Aug 2025 – present
- Developed a LLM-based method to deal with shortcut problem in general NLP model
- Improved the algorithm's stability, running speed and increased model inference accuracy from 0.8 to 0.92

Google Inc. Shanghai, China
Software Engineering Intern – AI-powered Accessibility Validation Jun 2025 – Aug 2025
- Leveraged Gemini to detect and validate end-to-end real accessibility issue cases and achieved over 0.95 recall rate.
- Engineered and automated an issue fetcher to enhance efficiency of internal issue collection pipelines.
- Deployed and integrated 12 accessibility criteria into production and reviewed the performance of 12 criteria.

Carnegie Mellon University Pittsburgh, PA
Research Assistant, *Supervised by Prof. Carl Kingsford* Dec 2024 – May 2025
- Implemented a RNA MFE and CAI parallel computing software, adapting different species codon database.
- Investigated and developed Minimum Free Energy predictor based on mRNA features includes secondary structure.
- Contributed to the model and scoring module for controllable mRNA sequence design, leading to the ARCADE paper.

University College London London, UK
Research Assistant, *Supervised by Prof. Christine Orengo* Jun 2023 – Sep 2023
- Trained and optimized a deep learning model for protein design in Python by redesigning model's structure and incorporating the SAM algorithm.
- Augmented datasets and extracted protein embeddings using multiple advanced protein language models.
- Applied statistical analyses to identify critical issues in deep-learning model training dynamics.

Zhejiang University-University of Edinburgh Institute Haining, China
Research Assistant, *Supervised by Prof. Chaochen Wang and Prof. Xue Jing* Jun 2022 – Jun 2023
- Confirmed role of Med12 to histones by performing integrated statistical analysis on multi-omics datasets.
- Applied statistical analyses and visualized downstream analysis results in R.
- Published original research as a Co-First Author in the high-impact journal, Gut (2024)

PROJECT EXPERIENCE

Protein function prediction by utilizing protein language model and GNN Oct 2023 – May 2024
Undergraduate Thesis Research, *Supervised by Prof. Christine Orengo*
- Developed and optimized an ensemble model (a protein Language Model and Graph Neural Network-based model) by using contrastive learning and hyperparameter tuning, achieving better F1 core compared to the state-of-art models.
- Established two baseline methods (MLP and Transformer) in PyTorch to rigorously benchmark the performance of the proposed ensemble architecture

Modeling changes in protein surface exposure April 2023 – May 2023

Computational Biology and System Biology Course Project, ZJE

- Defined mutations on VSG2 in silico and rebuilt the variants' protein structures using AlphaFold2
- Analysis protein surface changes caused by residue size and hydrophilia changes

SKILLS

- **Programming Language:** Python, R, Java, Linux, MySQL
- **AI/ML library:** Pytorch, TensorFlow, Keras, Scikit-learn
- **Bioinformatics:** AlphaFold2, ESM, Seurat, PyMOL, Biopython
- **Language skills:** English (fluent), Mandarin(native)

Personal Links

GitHub: https://github.com/ShijieTang
LinkedIn: http://www.linkedin.com/in/shijietang