

# Yu-Juan Zhang

Email: yujuan.zhang418@gmail.com; 41437909@qq.com

ORCID: 0000-0001-6361-0840

Research focus: protein language models,  
interpretable ML, graph AI, genomics, proteomics,  
multi-omics

Google Scholar:  
[scholar.google.com/citations?user=CdA1HhMAAAAJ&hl=en](https://scholar.google.com/citations?user=CdA1HhMAAAAJ&hl=en)

GitHub: [github.com/yujuan-zhang](https://github.com/yujuan-zhang)

Kaggle: [kaggle.com/yujuanzhangai](https://kaggle.com/yujuanzhangai)

## Appointments

Professor, College of Life Science, Chongqing Normal  
University, China

Dec 2017 - Jan 2026

Associate Professor, Chongqing Normal University,  
China

Dec 2012 - Nov 2017

Assistant Professor, Chongqing Normal University,  
China

Jul 2010 - Nov 2012

## Education

PhD, Cell Biology, Graduate School of the Chinese  
Academy of Sciences, China

Sep 2004 - Jul 2010

BS, Life Science, College of Life Sciences, Shaanxi  
Normal University, China

Sep 2000 - Jul 2004

## Research Interests

- Protein language models and interpretable representation learning (ESM2)
- Graph neural networks for biomedical link prediction (drug-disease association; drug repositioning)
- Multi-omics, Cancer Biomarker, Comparative genomics, Molecular evolution

## Technical Skills

- **Programming:** Python, R; Git; Unix/Linux; Shell
- **Machine Learning / Deep Learning:** scikit-learn, PyTorch; Transformers, GNNs, VAEs; contrastive learning
- **Data Analysis & Visualization:** pandas, NumPy; data cleaning & EDA; Matplotlib; Jupyter
- **Statistics & Experimentation:** hypothesis testing, confidence intervals, A/B testing, experimental design
- **Interpretability:** SHAP/feature attribution; embedding and attention analysis
- **Bioinformatics:** multi-omics analysis; NGS (WGS resequencing, RNA-seq); variant calling; GO/KEGG enrichment; pathway/network analysis; PubMed/DrugBank/PubChem mining; PyMOL visualization

## Teaching & Mentoring

Bioinformatics, Genomics and proteomics

## Selected Projects

### Interpretable Protein LLMs for ubiquitination prediction

- Built an ESM2-centered pipeline linking representation extraction, dimensionality reduction, and prediction.
- Developed interpretable models using embedding-based interpretation and attention-based analysis; contributed to code, data processing, and manuscripts.

### Graph AI for Predicting Drug-Disease Links (Drug Repositioning)

- Built a multi-view heterogeneous GNN to integrate noisy similarity networks and meta-paths for drug-disease association prediction.
- Improved robustness via contrastive learning; led implementation, benchmarking, and manuscript preparation.

### Multi-Omics for Cancer Biomarker Discovery

- Performed integrative analysis of WGS, RNA-seq, antibody-based protein profiling, and clinical metadata to support biomarker discovery and cross-cohort assessment; contributed across data preprocessing, QC, statistical analysis, and result interpretation.

### Integrative Multi-Omics in the Non-model Malaria Vector (*Anopheles sinensis*)

- Generated a de novo genome assembly and analyzed gene families associated with hematophagy and disease transmission.
- Integrated transcriptome, microRNA, metagenome, and resequencing data from pyrethroid-susceptible/resistant populations to infer resistance mechanisms (variants, selection, regulation, temporal expression); contributed end-to-end.

## Selected Publications (\* Corresponding author)

1. Jinzhou Wu, Donglin He, Xin Li, Rui Wang, Yajuan Zhang\*. A multi-view collaborative heterogeneous graph neural network with semantic- and relation-aware for drug-disease association prediction. Engineering Applications of Artificial Intelligence. 2026. DOI: 10.1016/j.engappai.2025.113217.
2. Yawen Sun, Rui Wang, Zeyu Luo, Lejia Tan, Junhao Liu, Ruimeng Li, Dongqing Wei, Yu-Juan Zhang\*. ESM2\_AMP: An Interpretable Framework for Protein-Protein Interaction Prediction and Biological Mechanism Discovery. Briefings in Bioinformatics. 2025. DOI: 10.1093/bib/bbaf434.
3. Zeyu Luo, Rui Wang, Yawen Sun, Junhao Liu, Zongqing Chen, Yu-Juan Zhang\*. Interpretable Feature Extraction and Dimensionality Reduction in ESM2 for Protein Localization Prediction. Briefings in Bioinformatics. 2024. DOI: 10.1093/bib/bbad534.
4. Junhao Liu, Zeyu Luo, Rui Wang, Xin Li, Yawen Sun, Zongqing Chen, Yu-Juan Zhang\*. EUP: Enhanced Cross-species Prediction of Ubiquitination Sites via a Conditional Variational Autoencoder Network Based on ESM2. PLOS Computational Biology. 2025 Jul 16;21(7):e1013268. DOI: 10.1371/journal.pcbi.1013268.

5. Yu-Juan Zhang, Zeyu Luo, Yawen Sun, Junhao Liu, Zongqing Chen\*. From Beasts to Bytes: Revolutionizing Zoological Research with Artificial Intelligence. *Zoological Research*. 2023;44(6):1115-1131.
6. Yu-Juan Zhang, Yang Lan, Bin Chen\*. ASDB: A comprehensive omics database for *Anopheles sinensis*. *Genomics*. 2021;113(3):976-982.
7. Xiaoyan Zuo, Bo Li, Chengxu Zhu, Zhengwen Yan, Li M, Xinyi Wang, Yu-Juan Zhang\*. Stoichiogenomics reveal oxygen usage bias, key proteins and pathways associated with stomach cancer. *Scientific Reports*. 2019 Aug 5;9(1):11344.
8. Yu-Juan Zhang, Chengxu Zhu, Yiran Ding, Zhengwen Yan, GongHua, Yang Lan, Jian-Fan Wen\*, Bin Chen\*. Subcellular stoichiogenomics reveal cell evolution and electrostatic interaction mechanisms in cytoskeleton. *BMC Genomics*. 2018;19(1):469.

## Books

- Yu-Juan Zhang, Boqi Li, Baoli Qiu. *Applications of Artificial Intelligence in Life Sciences*. Science Press, 2026. (In press).
- Yu-Juan Zhang and Wen JF. Proteobacteria and the endosymbiotic origin of mitochondrion. In: Maria L. Sezenna (ed.). *Proteobacteria: Phylogeny, Metabolic Diversity and Ecological Effects*. Nova Science Publishers, 2011, pp. 57-71.

## Patents and Software Copyrights

- Intelligent Assisted Ubiquitinated Protein Detection Platform V1.0 (Software Copyright). 2025. Registration No. 2025SR0582782. National Copyright Administration of the PRC.
- A Computational Method for Basic Stoichiometric Genomic Analysis (Invention Patent). Application No. 202110905333.9.
- A Joint Analysis Method for Stoichiometric Genomes and Genome Annotation Files (Invention Patent). Application No. 202110906004.6.
- A Batch Analysis Method for Multi-species Stoichiometric Genomes (Invention Patent). Application No. 2021109053.
- A High-throughput Method for Multi-species Stoichiometric Proteome Analysis (Invention Patent). Application No. 202210259556X.
- A High-throughput Method for Multi-species Stoichiometric Transcriptome Analysis (Invention Patent). Application No. 202210159540.

## Certifications

- Machine Learning Scientist in Python, Dec 2025, DataCamp.
- Associate Data Scientist in Python, Nov 2025, DataCamp.