Yujuan Zhang, PhD

Bioinformatics Scientist | NGS Multi-omics | AI/ML | Precision Medicine

Email: yujuan.zhang418@gmail.com

GitHub:
https://github.com/yujuan-zhang

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

Personal Website:
https://yujuan-zhang.github.io/

LinkedIn:
https://www.linkedin.com/in/yujuan-zhang-0a86b757/

Summary

Bioinformatics Scientist with ~20 years of experience applying quantitative and computational methods in the life sciences. I build reproducible pipelines spanning QC, variant annotation, drug mapping, and deployment-ready predictive models, and execute data strategies that support innovation in biotech and diagnostics. With a PhD in Bioinformatics, 15 years in NGS-driven multi-omics, 50+ peer-reviewed publications and patents, I am adept at driving R&D efforts from concept to completion. Professor-level leadership and deep expertise in AI/ML, cloud technologies, and NGS, together with active contributions via GitHub/Kaggle and extensive training and mentorship experience, have strengthened my approach to data integration and analysis, enabling impactful advancements in precision medicine while supporting a team's digital transformation goals.

Appointments

- **Chongqing Normal University, College of Life Science (China)**
  Professor (Level 3), Bioinformatics & Computational Biology | Jul 2023 – Jan 2026
  Professor (Level 4), Bioinformatics & Computational Biology | Dec 2017 – Jun 2023
  Associate Professor, Bioinformatics & Computational Biology | Dec 2012 – Nov 2017
  Assistant Professor, Bioinformatics & Computational Biology | Jul 2010 – Nov 2012
- **The University of Oklahoma, Department of Chemistry and Biochemistry**
  Visiting Scientist | Aug 2014 – Jul 2015

Education

- Chinese Academy of Sciences, China — PhD, Bioinformatics | Sep 2004 – Jul 2010
- Shaanxi Normal University, China — BS, Life Science | Sep 2000 – Jul 2004

Research Interests

- NGS, Multi-omics, Genomics, Biomarker, Precision diagnosis, pipeline development
- Protein language model–based functional and mechanistic inference
- Drug-protein-disease association, Drug repositioning

Core Skills

- **End-to-End Data/AI (E2E)**: data ingestion (ETL/ELT, Spark); SQL databases (PostgreSQL); data modeling & metrics (dbt); workflow orchestration (Airflow, Nextflow); MLOps (Docker/Compose; MLflow; CI/CD with GitHub Actions; deployment on AWS ECS/Fargate; monitoring with CloudWatch); web apps/dashboards (Streamlit); AWS data & infrastructure (S3, RDS)
- **Programming**: Python, R; Bash/Shell (Linux/Unix)
- **Machine Learning / Deep Learning**: scikit-learn, PyTorch, TensorFlow/Keras; classical ML (RF, XGBoost); deep learning (Transformers/PLMs, GNNs, VAEs); evaluation (stratified CV, AUROC, PR-AUC); interpretability (SHAP, Integrated Gradients); reproducible ML workflows
- **Data Analysis & Visualization**: pandas, NumPy; data cleaning & EDA; Matplotlib/Seaborn; Jupyter
- **Statistics & Experimentation**: hypothesis testing, confidence intervals, A/B testing, experimental design
- **Biology**: NGS (WGS/WES/RNA-seq/scRNA-seq); knowledgebase mining (NCBI, DrugBank, PubChem); pathway/network analysis; PyMOL; Population Genetics; Molecular Evolution; Microbiology; Molecular Biology
- **Genomics / Precision Medicine:** variant calling & annotation; cross-cohort integration (TCGA, CPTAC, CCLE, GTEx); multi-omics harmonization & feature engineering; tumor subtyping & risk stratification; survival modeling; AI-enabled clinical workflows

Mentorship & Training

- **Undergraduate:** Bioinformatics
- **Master's:** Advanced Bioinformatics, R for Statistical Computing & Graphics, Comparative Genomics, Proteomics & Genomics
- **PhD:** Advanced Bioinformatics

Selected Projects

LUAD Precision Oncology Platform (Streamlit, Docker, cloud-deployed):

- Built an end-to-end multi-omics analysis and clinician-facing reporting interface for the TCGA-LUAD cohort (n=560), integrating clinical metadata, MAF-based somatic variant profiling, bulk RNA-seq quantification, single-cell transcriptomics, ESM-based mutation-aware protein embeddings for predicting the deleteriousness of uncharacterized mutations, and biomarker/therapy knowledge mapping (OncoKB).
- Implemented scalable genomic analytical pipelines for large-scale cohort processing, including data cleaning, QC, and EDA, and delivered a unified integration layer spanning genomic features, patient covariates, and functional evidence to enable cross-omics interpretation and cohort-level prioritization.
- Developed an AI-enabled clinical decision support (CDS) workflow by producing structured, report-style interactive pages (variant summaries, actionability labels, evidence-linked therapy annotations) in Streamlit; containerized the platform with Docker for reproducible cloud deployment.

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 biomedical link prediction (Drug Repositioning).

- Built a multi-view heterogeneous graph neural networks (GNN) to integrate noisy similarity networks and meta-paths for drug-protein-disease association prediction.
- Led implementation, benchmarking, and manuscript preparation.

Selected Publications (* Corresponding author)

1. Jinzhou Wu, Donglin He, Xin Li, Rui Wang, **Yu-Juan 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.

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.

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.

Honor and Awards

- First Prize, National Business Science and Technology Progress Award; China Federation of Commerce; Dec 2022.
- Outstanding Young Scientist Award, World Conference on Biomedicine and Artificial Intelligence (WBAI), 2024.
- Leading Woman in Science and Technology Innovation, Shapingba District, Chongqing, 2023.
- Silver Award (University Track), Ascend AI Innovation Competition 2024 Chongqing Regional Finals; Project: U-Net-based intelligent single-cell detection framework; Awarding body: China AI Industry Development Alliance; Oct 2024; Role: First Supervisor.
- Bronze Award (Application Track), Ascend AI Innovation Competition 2023 Chongqing Regional Finals; Project: Applications of BERT language models in biological sequences and gene expression profiles; Awarding body: China AI Industry Development Alliance; Oct 2023; Role: First Supervisor.
- Third Prize, Chongqing Scientific and Technological Awards, China; Jun 2015.
- Zhuli Yuehua Scholarship; Chinese Academy of Sciences; May 2010.
- First Prize, Scientific Paper Award; Yunnan Province, China; Jan 2010.
- Outstanding Article Award; State Key Laboratory of Genetic Resources and Evolution, China; Jun 2009.
- First-Class Scholarship; Shaanxi Normal University; Jan 2003.

Grants

- 2022 Natural Science Foundation Project of Chongqing (CSTB2022NSCQ-MSX0650), Stoichioomic study of the adaptive evolution mechanism of biomolecular in altered stomach environment and application. China, PI.
- 2020 Open Project of National Key Laboratory of Microbiology (MML-202005), Research of adaptive evolution of sulfur bacteria based on stoichioomics, China, PI.
- 2018 National Natural Science Foundation of China (Grant No. 3187090549), A study of cellular compartments and construction of integrated analysis platform based on stoichiogenomics, China, PI.
- 2018 Natural Science Foundation Project of Chongqing (CSTC2018jcyjA2487), China, PI.
- 2016 Science and Technology Project of Chongqing Municipal Education Commission (No. KJ1600304), China, PI.
- 2013 Natural Science Foundation Project of Chongqing (CSTC2013JCYJA00013), China, PI.

- 2013 Science and Technology Project of Chongqing Municipal Education Commission (No. KJ130629), China, PI.
- 2012 National Natural Science Foundation of China (Grant No. 31200947), Protein transport system of secondary plastids (chloroplasts) in two diatoms species and its evolution research, China, PI.
- 2011 Open Project of State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Science (Grant No. GREKF11-09), Comparative genomic analysis of protein transport in secondary plastids (chloroplasts), China, PI.
- 2010 Doctoral Start Fund in Chongqing Normal University (Grant No. 11XLB001), China, PI.

Professional Service

- Session Chair, World Conference on Biomedicine and Artificial Intelligence (WBAI), 2024-2025.
- Member, Genomics Professional Committee, Entomological Society of China, 2022-2027.
- Editorial Board Member, Interdisciplinary Sciences: Computational Life Sciences, 2024-2026.
- Executive Council Member, Chongqing Society of Bioinformatics.
- Panel Reviewer/Expert, National Natural Science Foundation of China (NSFC).

Certifications

- Machine Learning Scientist in Python, Dec 2025, DataCamp.
- Associate Data Scientist in Python, Nov 2025, DataCamp.
- Kaggle competition participation (Rank ~3,000/15,000 teams).
- Editorial Board Member, <Interdisciplinary Sciences: Computational Life Sciences>