

## EDUCATION BACKGROUND

Columbia University | Mailman School of Public Health

New York, NY

Master of Science in **Biostatistics**

May 2026 (expected)

**Coursework:** Data Science, Statistical Inference, Biostatistical Methods, Probability, Principles of Epidemiology

Fudan University | School of Life Sciences

Shanghai, CN

Bachelor of Science in **Biological Sciences**

Jul. 2024

**Academic Honors:** 3<sup>rd</sup> Prize of Fudan University Merit Scholarship for twice in Oct. 2021 and Oct. 2022

## PUBLICATION LIST

- **Preparation:** Minghan Li, **Jianming Wang**, Weidong Tian, SWORD: Single-Cell Reference with Integrated Framework for Deconvolution
- **Submitted:** Yucan Li, Xinming Xu, Yi Zheng, Rui Li, Xin Zhang, Jiacheng Wang, Ningxin Gao, **Jianming Wang**, Yawen Wang, Jialin Li, Jincheng Li, Danke Wang, Zhenqiu Liu, Mei Cui, Yanfeng Jiang, Yingzhe Wang, Chen Suo, Tiejun Zhang, Kelin Xu, Xingdong Chen, Synergistic and heterogeneous aging: insights from composite phenotypes and multiple organ systems aging clocks, *Advanced Science*.
- **Published:** **Jianming Wang**, Research of the Influencing Factors of Mental Health Status of Students in Leeds-based on the 2019-20 Student Health and Well-being Survey., *Journal of Education, Humanities and Social Sciences*, Volume 17 (2023).

## PROFESSIONAL SKILL

### OPERATION TOOLS

- R, Python, SQL, SPSS, Linux

### ANALYTICAL METHODS

- Network Analysis, Correlation Analysis, Significance Test, ANOVA
- Cross Validation, Bootstrap, Random Forest, Lasso, Ridge Regression, KNN, Linear/Logistic Regression Models
- PCA, Cluster Analysis, Developmental Trajectory Analysis, GSEA, Differential Expression Analysis

## RESEARCH EXPERIENCE

### Exploration of A-to-I Editing Sites Based on SNP Allele Frequencies and PhyloP Scores

*Core Participant, supervised by Prof. Weidong Tian, Fudan University*

Dec. 2023 – May 2024

- Conducted an in-depth analysis of A-to-I RNA editing sites that overlap with Single Nucleotide Polymorphisms (SNPs), focusing on their distribution, conservation, and functional characteristics using statistical and bioinformatics approaches.
- Utilized public databases such as REDportal, dbSNP, UCSC, and NCBI SRA to gather comprehensive datasets of A-to-I editing sites and SNPs, employing rigorous data cleaning and preprocessing techniques.
- Performed Pearson's Chi-square test to reveal the conservation features of SNP editing sites within human populations and throughout evolution.
- Employed the SPRINT tool to identify RNA editing sites from RNA-seq data, contributing to a better understanding of the distribution and characteristics of SNP editing sites.
- Executed gene enrichment analysis to uncover associations between SNP editing sites and cellular morphogenesis, neural activity, as well as potential links to cancer and various mental disorders.
- Authored a comprehensive thesis detailing the research methodology, findings, and implications, highlighting the significance of SNP editing sites in RNA editing research and their potential impact on biological research, drug development, and clinical treatment.

### SWORD: Single-Cell Reference with Integrated Framework for Deconvolution

*Assistant Participant, supervised by Prof. Weidong Tian, Fudan University*

Jul. 2023 – Sep. 2023

- Calculated the expression profile of the cell type for each reference patient and the Pearson correlation coefficient between different reference patients, completed the screening and merging of all reference patient single-cell atlas data
- Applied the existing deconvolution algorithm (e. g. LM, RLM, PCR, etc.) to deconvolute the expression profile data to obtain several different deconvolution results
- Performed PCA dimension reduction of cell types to complete preliminary cluster analysis, and deleted outliers
- Developed the SWORD, a deconvolution framework for integrating multi-reference datasets, and tested on the pseudo-bulk data and real bulk data set respectively to conclude that the excellent performance of SWORD

- Evaluated the results with SWORD based on three indicators of minimum spanning tree (MST), correlation coefficient within the group (ICC) and standard deviation (SD) within and between group between different deconvolution results
- Employed Entropy Weight Method to calculate to conclude that the higher the similarity between different results, the higher the reliability and accuracy of the integrated results
- Completed the paper titled *SWORD: Single-Cell Reference with Integrated Framework for Deconvolution*

**Multi-organ Biological Age Estimation based on Multiphenotypic Data**

Mar. 2023 – Jun. 2023

**Core Participant, supervised by Dr. Kelin Xu, Fudan University**

- Utilized R to preprocess and classify the indicators data from Taizhou Imaging Study (TIS), including blood biochemistry, bone density, blood vessel elasticity, etc.
- Performed Network Analysis to select appropriate variables, including the correlation calculation between each indicator as well as indicators and ages, and network diagram drawing
- Calculated and visualized the differences in the impact of selected variables on biological age among gender groups in the TIS
- Applied the KDM algorithm to obtain the predicted physiological age and verify it with survival data
- Conducted external data verification and project promotion for the above programs based on the public data from NHANES
- Completed the paper titled *Multimodal Aging Clock in the Cohort of Older Adults in Rural China* and plan to further refine selection methods and predictive models

**Validation and Fine Classification of T-Cell Subsets According to Marker Genes based on Single-cell Data**

Aug. 2022 – Jan. 2023

**Independent Researcher, supervised by Prof. Weidong Tian, Fudan University**

- Summarized 100+ pieces of marker genes through literature review, and used R for data cleaning and cluster analysis
- Verified the marker genes through public data sets such as PBMC to confirm the validity of the gene set
- Used R to visualize and analyze the expression of different marker genes in the data set with heat and dot maps to show the differences in expression levels, and conducted statistical tests to verify their significance
- Classified and annotated marker genes from perspectives of surface antigens, developmental stages and executive functions, and constructed a T cell reference containing 290,588 single-cell data information
- Applied Monocle and VECTOR to conduct developmental trajectory analysis on exhausted T cell data, and detected cell marker genes at different developmental stages, thereby verifying the validity of the collected exhausted T cell marker genes at each developmental stage
- Proposed a finely classify method for T cell subtypes from surface antigens to developmental stages and functions, and applied real single cell data sets to verify the feasibility and accuracy of this method
- Plan to build a T cell reference based on multiple classification dimensions and develop tools for multidimensional labeling of T cells

**ACADEMIC HIGHLIGHT****Exploring the Effect of Vaccination on Reducing Severe Cases based on Worldly COVID-19 Vaccination Data**

Feb. – Apr. 2023

- Applied R to collect data on indicators such as the total number of infected people, mortality rates, and vaccination rates in the global epidemic on Kaggle, and completed data preprocessing and visualization
- Constructed a correlation model based on linear regression and logistic regression to predict the trend of severe disease rate with vaccination rate

**Gene Function Verification of Drosophila**

Nov. – Dec. 2022

- Designed the experimental steps of cloning the sgRNA expression cassette, obtained the appropriate sgRNA for targeting the target gene sequence, and edited the RNA at a fixed point
- Used CRISPR-Cas9 gene editing technology to construct Drosophila mutants to verify gene function

**Investigation and Analysis of Human Genetic Traits**

Oct. – Nov. 2022

- Explored dimple traits from the aspects of trait ratio, gender and geographical distribution of traits
- Constructed a pedigree based on the survey results, performed correlation analysis with R and SPSS to better acquire genetic traits