# Cai, Yilin (Scarlett)

caiyilin1205@g.ucla.edu | 310-448-3577

## EDUCATION

**University of California, Los Angeles** **Sep. 2021—Present**

*B.S. in Computational & Systems Biology, Minor in Statistics & Data Science* GPA 3.92/4.0 Los Angeles, CA

**Relevant Coursework:** Computational & Systems Biology, Bioinformatics, Epidemiology, Multivariable Calculus, Biological Modeling: Mathematical and Computational Approaches, Computer Science

**Honors:** Dean's Honors List (2024 Spring/Winter, 2023 Spring/Winter, 2022 Fall/Spring/Winter, 2021 Fall)

## PROFESSIONAL EXPERIENCES

**Origincell Bank Co., Ltd** **Jul. 2024—Aug. 2024**

*Intern* Shanghai, China

- Assisted in the creation of marketing materials a stem cell storage and immunotherapy, focusing on translating complex clinical information into accessible formats for public engagement.
- Supported the production and editing of promotional videos, enhancing the company's outreach to potential clients and healthcare partners.

**Livzon Pharmaceutical Group** **Jul. 2023—Sep. 2023**

*Intern* Shanghai, China

- Assisted the Medical Department in overseeing hormone drug trials for breast and prostate cancer, including tasks such as monitoring blood sample collection and preservation, auditing participant records
- Utilized SPSS for data analysis to conduct bioequivalence studies, ensuring compliance with FDA standards
- Supported colleagues with literature review, translation, and data analysis

## RESEARCH & PROJECTS

**Oppenheimer Center for Neurobiology of Stress and Resilience, Gupta Lab** **Sep. 2022—Jun. 2024**

*Research Assistant* Los Angles, CA

- Engaged in a postdoc research project regarding the association between stress/discrimination exposure and functional brain activity (task-fMRI) on adolescent obesity (BMI) with genetic data.
- Performed data cleaning, visualization, and statistical analysis using R, Python, Excel, and MATLAB of the ABCD data repository
- Utilized self-reported data and fMRI information from a large database to identify patterns in social factors, demographic characteristics, and brain activity that may influence BMI among adolescents.
- Applied Partially Square (PLS) analysis and other statistical methods to explore potential correlations between social segregation measures, amygdala connectivity, and BMI

**Efficient Pseudoalignment Implementation for Enhanced RNA-Seq Data Analysis and Transcript Quantification** **Apr. 2024—May 2024**

- Implemented a pseudoalignment algorithm for RNA-seq data analysis, efficiently mapping reads to a reference transcriptome using shared k-mers, enabling precise transcript quantification across biological samples.
- Developed functions to read and process RNA-seq data in FASTA format, generate k-mers, and construct a k-mer index, improving the computational efficiency and accuracy of read mapping.
- Analyzed equivalence class distributions, visualizing unique and ambiguous transcript alignments, and identified areas for optimization to enhance alignment precision and reduce computational demands.

**Improving Gradient Boosting in SCING to Aid GRN Inference from Single Cell Transcriptomics** **Jan. 2024—Mar. 2024**

- Enhanced the SCING framework for gene regulatory network inference by comparing Gradient Boosting Decision Tree (GBDT) with alternative models, XGBoost and LightGBM, achieving improved predictive accuracy for single-cell transcriptomic data.
- Implemented model modifications including permutation importance to assess feature contributions, optimizing

SCING's ability to handle complex gene interactions and improve regulatory pattern recognition.

- Leveraged the Hoffman2 Cluster for computational support, achieving efficient processing and visualization of gene networks, and demonstrated XGBoost's superior AUROC score over GBDT in predictive tasks

## EXTRACURRICULAR ACTIVITIES

**In Transcription. @UCLA** *Core Member* **Sep. 2022—Present**

- Participated in weekly meetings and learned about various topics in biomedical research and industry; presented personal experiences on participating research and joining laboratories
- Engaged in volunteer activities in local communities.

**UCLA Statistics Club** **Member** **Sep. 2021—Jun. 2022**

- Participated in the Python workshops; conducted several data analysis tasks and projects using R.

## LANGUAGES & SKILLS

**Languages:** Mandarion (Native), English (Fluent), French (Beginner)

**Technical Skills:** Microsoft Suite, Python, C++, R, Pandas, Canva, SPSS, MATLAB