Shang-Yang (Sam) Chen

sychen9584@gmail.com || Boston, MA

sychen9584.github.io || US Permanent Resident – No sponsorship required

PROFESSIONAL SUMMARY

- Ph.D.-trained computational biologist specializing in **single-cell analysis, multi-omics integration,** and **scalable pipeline development.**
- Experienced in building production-ready tools and pipelines for **transcriptomic** and **epigenomic** data, with applications in **gene therapy, oncology,** and **immunology**.

EXPERIENCE

SCIENTIST I/II, COMPUTATIONAL SCIENCE | DYNO THERAPEUTICS | JUNE 2022 - AUGUST 2024

- Designed and deployed production-grade Python pipelines for automated processing, QC, and analysis of single-nucleus RNA-seq (snRNA-seq) data, enabling scalable analysis of **1.5M+ nuclei** per dataset and reducing analysis time by **75%**.
- Served as primary analysis lead for snRNA-seq data across ocular, brain, and liver programs, supporting evaluation and validation of AI-designed AAV gene therapy vectors, contributing to **two product launches** and a **licensing agreement**.
- Performed integrative analysis of bulk NGS datasets from non-human primates, mice, and in vitro assays to guide design and selection of high-performing AAV variants.
- Collaborated with biologists, engineers, and leadership to optimize experimental design, scale data pipelines, and support IP filings and conference presentations.

Bioinformatics Intern | GENENTECH, ONCOLOGY INFORMATICS | OCTOBER 2021 – APRIL 2022

- Developed the R package *Epiregulon*, published on **Bioconductor**, enabling inference of transcription factor (TF) activity in single cells via gene regulatory network construction; integrated scATAC-seq, scRNA-seq, and public ChIP-seq data to support therapeutic discovery targeting TFs in oncology.
- Designed and implemented a benchmarking framework to evaluate *Epiregulon* against reference TF-target gene databases, ensuring rigorous model validation and robust performance assessment.
- Analyzed and interpreted scATAC-seq datasets to identify transcriptional regulatory mechanisms in breast cancer subtypes, supporting early oncology discovery efforts.

Graduate Researcher | NORTHWESTERN UNIVERSITY | AUGUST 2016 - MAY 2022

- Led integrative, multi-omics analyses (scRNA-seq, bulk RNA-seq, CITE-seq, ATAC-seq) to characterize immune cell heterogeneity and inflammatory pathways in rheumatoid arthritis and liver cholestasis.
- Designed and implemented custom R pipelines for scRNA-seq and CITE-seq, enabling high-resolution profiling of immune cell diversity.
- Developed and deployed *MAGNET*, a web-based tool for functional enrichment analysis using curated immunology gene sets, adopted by research teams for hypothesis generation and data interpretation.
- Awarded the American Heart Association (AHA) predoctoral fellowship in recognition of research excellence in computational immunology.

EXPERTISE

- **Core Skills:** Bioinformatics · Single Cell · Transcriptomics · Epigenomics · Immunology · Multi-omics Integration · Pipeline Development · Statistical Modeling · ML/AI · Cloud Computing · Data Visualization · Web Application Development
- **Programming Languages:** Python · R · SQL · Java · UNIX/Linux
- **Frameworks & Libraries:** Seurat · Scanpy · Nextflow · STAR · Samtools · DESeq2 · BioPython · FastQC · squidpy · Pandas · NumPy · Scikit-learn · PyTorch · AWS · GCP · Docker · Git · ggplot2 · Matplotlib · Tidyverse · Django

EDUCATION

Ph. D. | Biomedical Informatics | Northwestern University | 2016-2022

BSc | Biochemistry | Stony Brook University | 2012 - 2016