

Shan A. Sabri

COMPUTATIONAL BIOLOGIST · AI/ML · MULTI-OMICS

630-201-1911 | ShanASabri@gmail.com | ShanSabri.github.io | ShanSabri | ShanSabri | [Publications](#)

Summary

Computational biology leader with end-to-end expertise in AI/ML methods for multi-modal biological data. Developing novel approaches integrating foundation models, computer vision, and single-cell genomics. Proven track record building scalable data pipelines and agentic workflows for automated analysis and reporting. Demonstrated success in clinical translation and cross-functional collaboration.

Education

University of California, Los Angeles

Ph.D. IN BIOINFORMATICS, EMPHASIS MACHINE LEARNING ([DISSERTATION](#))

Los Angeles, CA

June 2020

Johns Hopkins University

M.Sc. IN BIOINFORMATICS

Baltimore, MD

Dec. 2013

Loyola University Chicago

B.Sc. IN BIOINFORMATICS AND MATHEMATICS, MINOR BIOSTATISTICS

Chicago, IL

May 2012

Experience

Cellanome

San Diego, CA

Sr. STAFF SCIENTIST → PRINCIPAL SCIENTIST, COMPUTATIONAL BIOLOGY

Feb. 2024 – Present

- Manage a team of scientists; lead computational strategy for 15+ external collaborations with industry and academic partners ([posters](#))
- Drive enterprise LLM adoption; architect agentic AI workflows for biological data interpretation, research automation, and developer tooling
- Lead end-to-end evaluation of DINOv2 morphology features from image inputs; validate biological interpretability for phenotype prediction
- Develop production-grade analytical software enabling end-to-end analysis of multi-modal single-cell data; establish documentation, vignettes, and CI/CD pipelines for reproducible workflows

Sana Biotechnology

Chicago, IL

LEAD COMPUTATIONAL BIOLOGIST, EMBEDDED

Jan. 2023 – Jan. 2024

- Developed computational screening pipeline integrating DepMap and scRNA-seq to identify 14 essential gene targets; findings enabled genome engineering of inducible safety switches for selective elimination of undifferentiated cells in iPSC therapeutics
- Architected scalable QC pipelines and interactive dashboards for multi-omics analysis and interpretation; established release assay workflows characterizing differentiation purity and conversion efficiency for beta islet cell therapy manufacturing ([clinical trial](#))

ArsenalBio

South San Francisco, CA

SCIENTIST II → SENIOR SCIENTIST I, COMPUTATIONAL BIOLOGY & MACHINE LEARNING

Aug. 2020 – Sept. 2022

- Identified GS94 safe harbor locus by surveying 300,000 genomic sites computationally, prioritizing 39 candidates for experimental validation; discovery directly enabled AB-1015 CAR-T program for ovarian cancer ([clinical trial](#))
- Built an interactive atlas of >1.2M single cells integrating transcriptomics and surface proteomics (130+ antibody CITE-seq panel) across 45 donors, 5 tissues, and 3 treatment conditions; enabled tumor microenvironment characterization for CAR-T target discovery

UCLA

Los Angeles, CA

PH.D. CANDIDATE, BIOINFORMATICS (ERNST & PLATH LABS)

Aug. 2014 – June 2020

- Profiled >200,000 single cells across three somatic cell types to characterize transcriptional dynamics during iPSC reprogramming; identified conserved gene network silencing mechanisms governing cell fate transitions (16 publications in *Cell* and *Nature* family journals)
- Genentech (Summer 2018): Integrated multi-omics data to characterize tumor cell plasticity and identify pathway drivers in metastatic CRC ([poster](#))

Skills

Machine Learning / AI

Foundation models (scGPT, GeneFormer); deep learning (PyTorch, TensorFlow); computer vision (DINOv2, YOLOv5, Mask R-CNN); agentic AI workflows

Computational Biology

Single-cell analysis (Seurat, Scanpy, scVI); multi-modal integration (scRNA-seq, scATAC-seq, CITE-seq); epigenomics (ATAC-seq, ChIP-seq, CUT&RUN); spatial transcriptomics; trajectory inference & RNA velocity; CRISPR screens & off-target analysis; biological databases (UniProt, GenBank, GEO, Ensembl, GO, KEGG, Reactome)

Programming

Python, R; Dash & Shiny; data systems (SQL, NoSQL); Nextflow, Snakemake; AWS/HPC; Docker; Git