

SAI MIRRA G

Bioinformatics & Data Science
Researcher

CONTACT

saimirraol@gmail.com
Chennai, India
github.com/smir9ra18
<http://www.linkedin.com/in/saimirrag99>

SKILLS

Python NumPy · Pandas · scikit-learn
R DESeq2 · ggplot2 · Bioconductor
MATLAB Analysis · Scripting
Bioinfo Scanpy · Biopython · BLAST
GEOparse · Galaxy · Ensembl
ML Regression · Classification · Clustering
Viz Matplotlib · Seaborn · ggplot2
VCS Git · GitHub · Snakemake

RESEARCH DOMAINS

- Transcriptomics & Omics
- Computational Neuroscience
- Systems Biology
- Aging & Longevity
- AI for Molecular Biology

CERTIFICATIONS

NPTEL Elite — Python for Data Science
IIT Madras · Jan–Feb 2025
MATLAB Onramp + Desktop Tools + Data Plots
MathWorks · Apr 2025

WORKSHOP

CRISPR & Gene Editing Tools
IISc, Bangalore

SOFT SKILLS

- Analytical Thinking
- Research Orientation
- Adaptability & Collaboration
- Critical Thinking

M.Sc. Bioinformatics & Data Science graduate with end-to-end experience building computational biology pipelines — from raw GEO datasets to trained models and interpretable visualisations. Core strengths in transcriptomics, single-cell analysis, and applied machine learning. Currently developing a Variational Autoencoder to model transcriptomic brain ageing trajectories in the human prefrontal cortex.

EDUCATION

M.Sc. Bioinformatics & Data Science

Sathyabama Institute of Science and Technology, Chennai 2024 – 2026

- Systems Biology · Computational Genomics · Machine Learning · Structural Bioinformatics · Biostatistics

B.Sc. Plant Biology & Plant Biotechnology

Meenakshi College for Women (Affiliated to University of Madras) 2021 – 2024

- Undergraduate research: Phytochemical and Antimicrobial Analysis of Delonix regia Floral Extract

PROJECTS

VAE-Based Modelling of Transcriptomic Brain Ageing — Human Prefrontal Cortex

Tools: Python · PyTorch · Scanpy · GEOparse · Matplotlib

- Designing a VAE to learn latent representations of bulk RNA-seq profiles from age-stratified prefrontal cortex samples.
- Modelling continuous ageing trajectories from high-dimensional gene expression data via dimensionality reduction.
- Identifying age-associated gene modules linked to neurodegeneration and longevity.

Single-Cell RNA-seq Workflow — Mouse Brain (GSE67835)

Tools: Python · Scanpy · GEOparse · Pandas · Matplotlib

- Built end-to-end scRNA-seq pipeline on adult mouse cortex and hippocampus; QC, normalisation, UMAP, Leiden clustering.
- Annotated major brain cell populations using known marker genes.

Differential Gene Expression Analysis

Tools: R (DESeq2, ggplot2) · Python (GEOparse, Seaborn)

- Applied DESeq2 on NCBI GEO RNA-seq data to identify up- and down-regulated genes between biological conditions.
- Visualised results via volcano plots and heatmaps.

Brain Age Prediction — Neuro-AI Demonstration

Tools: Python · scikit-learn · NumPy · Matplotlib

- Implemented Linear Regression and Random Forest to estimate brain age from simulated MRI-derived features.
- Evaluated with RMSE and R^2 ; conducted feature importance analysis.

PPI Network Construction & Protein Annotation

Tools: Python · NetworkX · STRING API · Biopython

- Built disease-related PPI network from STRING API; computed centrality metrics to identify hub proteins.
- Fetched UniProt sequences and extracted GO terms for functional annotation via Biopython.

EXPERIENCE

AI & ML Virtual Internship

YBI Foundation — Online (Remote) Apr 2025 · 2 weeks

- Applied supervised and unsupervised ML algorithms to structured datasets using Python and scikit-learn.
- Completed full pipeline: data preprocessing, feature selection, and model evaluation.