Santoshi Borra

945-268-3572 bsantoshi.d@gmail.com LinkedIn GitHub

Education

Indiana University, Bloomington (*Master of Science in Bioinformatics and Informatics*) Aug 2023 – May 2025
Sri Ramaswamy Memorial University (SRMU) (*Bachelors in Biotechnology, Genetics*) July 2019 – May 2023

Skills

Programming Languages: Python, R , SQL, Bash/UNIX, Perl
Machine Learning & AI: Supervised Learning (Classification, Regression), Neural Networks, Deep Learning, Model Evaluation, TensorFlow, PyTorch, Scikit-learn
Frameworks & Tools: Tidyverse, Bioconductor, R Shiny, Plotly, ggplot2, Matplotlib, Tableau, Pandas, Biopython, Snakemake, Nextflow, Conda, SLURM
Cloud Computing & DevOps: AWS (EC2, S3, SageMaker), Docker, Git/GitHub
Bioinformatics & Computational Biology: Next Generation Sequencing Analysis (RNA-seq, scRNA-seq, ChIP-seq), Metagenomics, Proteomics, Metabolomics, Comparative Genomics, GWAS, Computational Tissue Modeling, Biostatistics
Databases & Genomic Resources: BLAST, NCBI, UCSC Genome Browser, Ensembl, UniProt

Experience

Ourobio (*Bioinformatician*) June 2025—present

- Investigated catabolite repression mechanisms and transcriptional regulators (CRP, LacI) limiting mixed carbon source utilization (glucose, lactose) using FlexFlux and COBRApy. Developed a computational model integrating metabolic and regulatory networks to identify bottlenecks in PHA/indigo production pathways.
- Simulated gene knockouts overexpression strategies to alleviate carbon catabolite repression, enabling simultaneous substrate fermentation. Predicted optimized regulatory genetic architectures to enhance product yields from waste feedstocks. Validated model predictions against experimental data and proposed actionable genetic modifications for wet-lab strain improvement.

Indiana University (*Research Assistant*) May 2024—May 2025

- Investigated NK cell heterogeneity and transcriptional regulation in chronic myeloid leukemia using multi-omics approaches. Designed and executed an end-to-end scRNA-seq pipeline to analyze more than 300,000 immune cells from CML patient samples.
- Reconstructed context-specific gene regulatory networks using GENIE3's ensemble ML framework to identify master transcriptional regulators driving relapse pathogenesis; validated targets through TCR-seq clonotype expansion mapping linking cytotoxic dysfunction to NK cell exhaustion; proposed an IL-15–mediated reprogramming model with therapeutic implications.

Apex Institute (*Intern*) Jun 2021— Jan 2022

- Engineered integrated analysis of scRNA-seq (murine T/B cells) and bulk RNA-seq (germinal center B cells) to map lymphocyte activation trajectories during immune challenge. Identified novel Pdcd1+/Havcr2+ transcriptional programs driving T-cell dysfunction through cross-platform validation of scRNA-seq clusters against flow cytometry protein signatures.
- Reconstructed B-cell differentiation trajectories using RNA velocity and SCENIC TF regulons, revealing STAT3-mediated germinal center formation mechanisms. Automated Snakemake pipeline to harmonize 15+ public datasets (GEO/SRA), enabling discovery of conserved exhaustion markers across human systems.

Projects

Dynamic Analysis of Kinase-Phosphatase Interplay in Yeast Respiration Using Temporal Graphs

- Pioneered a temporal graph framework to study dynamic kinase-phosphatase interactions in yeast using gene expression data and static PPI networks. Applied CTMP-a graph based community detection with clustering to identify evolving protein interaction patterns.
- Built a customized GC-LSTM encoder-decoder pipeline for dynamic link prediction across metabolic cycles. Achieved 100% Hits@20 accuracy in predicting future protein-protein interactions under optimal training conditions.

Identification Of Selective Inhibitors For STAT6 Against Rheumatoid Arthritis *Molecular Docking*

- Studied and optimized selective STAT6 inhibitors using advanced computational techniques including free energy perturbation calculations and microsecond-scale molecular dynamics simulations to evaluate binding stability and predict selectivity profiles.
- Established an integrated structure-based drug discovery pipeline combining high-throughput virtual screening of 1M+ compounds, pharmacophore modeling and fragment-based design (MOE), and rigorous binding free energy validation. Resulted in three novel chemotypes with sub-120 nM predicted affinities.
- Led translational validation efforts, collaborating with medicinal chemists to synthesize top candidates and establish biophysical assays confirming nM-range binding. Advanced two compounds to hit-to-lead stage with favorable ADMET properties.

Genomic Analysis of Beetle and Nematode Species *PacBio HiFi, Hi-C, Synteny*

- De novo assembled genomes using PacBio HiFi reads and Hi-C chromatin capture data. Scaffolded using YAHs, achieving N50 more than 50Mb and BUSCO completeness. Optimized assembly parameters for repeat resolution and contiguity.
- Identified species-specific genomic regions through comparative analysis. Performed synteny mapping with Heterodera and C. elegans using RIdeogram, revealing conserved loci and evolutionary divergence. Designed a reproducible pipeline for assembly validation and annotation.