Santoshi Borra

+1 945-268-3572 | santoshiborra09@gmail.com | linkedin | github | portfolio

SUMMARY

Results-driven Bioinformatician with hands-on experience in spatial and single cell transcriptomics, systems biology modeling, and large-scale genomic analysis. Expertise in developing production-grade computational pipelines using Python, Nextflow, Docker, and HPC (SLURM) to analyze high-dimensional biological data. Delivered measurable impact in translational oncology, metabolic engineering, and drug discovery, including spatial tumor–immune profiling, predictive modeling, and high-throughput virtual screening. Known for strong cross-functional collaboration and the ability to bridge computational methods with experimental and clinical research needs.

PROFESSIONAL EXPERIENCE

Bioinformatician - Translational Research, **Singapore General Hospitals - Singhealth** **Dec 2025 - Present**

- Applied NanoString CosMx spatial transcriptomics with Seurat-based scRNA-seq reference mapping to characterize the tumor–immune interface in LSD1-inhibited TNBC across 100K+ spatial cells, enabling spatial program discovery.
- Built a qPCR genotype caller using ΔCq/ΔΔCq features, GMMs, and logistic calibration, achieving greater than 99% concordance and reducing manual genotype review time by 70% across high-throughput runs.
- Implemented run-level QC and deterministic no-call rules using control wells and replicate concordance, lowering false genotype assignments by 35% and improving cross-batch reproducibility.
- Partnered cross-functionally with clinicians and wet-lab teams to translate spatial and molecular insights into actionable oncology research outcomes.

Bioinformatician, **Ourobio** **Jun 2025 - Nov 2025**

- Engineered an end-to-end computational modeling framework using Python, Bash, and COBRApy to integrate constraint-based metabolic models with transcriptional regulatory networks, standardizing model construction and reducing manual iteration cycles by 60% across multiple engineered microbial strains.
- Deployed large-scale kinetic and perturbation simulations on HPC clusters using SLURM, executing thousands of gene knockout and overexpression scenarios per run, enabling systematic prioritization of high-impact genetic edits that were infeasible to test experimentally.
- Identified and ranked genetic modifications that increased predicted yields of industrial compounds (PHA, indigo) by greater than 90% relative to baseline models, directly informing experimental design choices and lowering wet-lab trial-and-error costs.
- Implemented model validation, sensitivity analysis, and parameter benchmarking workflows to assess flux stability and regulatory robustness across simulated conditions, increasing confidence in prioritized edits and reducing downstream experimental rework by 30%.
- Collaborated cross-functionally with metabolic engineers and experimental biologists to translate in silico predictions into testable hypotheses, improving model-to-experiment alignment and iteration speed.

Research Assistant, **Indiana University** **May 2024 - May 2025**

- Developed a scalable, Dockerized scRNA-seq analysis pipeline using Seurat and Linux, processing 300K+ immune and NK cells from CML patient cohorts, ensuring reproducibility, portability, and consistent performance across large clinical datasets.
- Performed differential expression, clustering, and pseudotime analysis to resolve relapse-associated cellular trajectories, identifying 9 novel biomarkers and disease-state transitions that improved biological signal detection by 25% over baseline workflows.
- Reconstructed context-specific gene regulatory networks using GENIE3 and SCENIC, identifying master regulators of relapse, and validated findings through TCR-seq clonotype analysis and CIBERSORTx immune deconvolution to strengthen causal inference.
- Worked cross-functionally with clinicians, immunologists, and computational scientists to propose an IL-15–mediated immune reprogramming model, shaping therapeutic hypotheses and follow-up studies.

Research Intern, **Apex Institute** **Jun 2021 - Jan 2022**

- Built a reproducible, Nextflow-based multi-omics workflow integrating scRNA-seq, bulk RNA-seq, and ATAC-seq, enabling joint analysis of transcriptional and chromatin accessibility data to map lymphocyte activation trajectories.
- Reconstructed B-cell differentiation dynamics using RNA velocity and SCENIC regulon analysis, improving temporal resolution of immune state transitions compared to static clustering approaches.
- Identified dysfunctional T-cell programs (Pdcd1+/Havcr2+) through cross-platform validation with ChIP-seq and CyTOF, uncovering STAT3-driven regulatory mechanisms underlying germinal center formation.
- Collaborated closely with experimental immunology teams to align computational insights with functional assays, increasing biological interpretability and translational relevance of results.

EDUCATION

Indiana University, Bloomington - *Master of Science*, *Bioinformatics and Informatics*

Sri Ramaswamy Memorial University, Hyderabad, India - *Bachelor of Science*, *Biotechnology, Genetics*

PROJECTS

Dynamic Analysis of Kinase Phosphatase Interplay in Yeast Respiration

- Engineered a temporal graph modeling framework integrating time-series gene expression, PPI networks, and CTMP-based link prediction, achieving 100% Hits@20 accuracy and >30% improvement over static models, identifying key kinase–phosphatase regulatory nodes and graph embeddings that accelerated respiration pathway mechanism discovery.

Identification of Selective STAT6 Inhibitors for Rheumatoid Arthritis

- Developed a high-throughput molecular docking and simulation pipeline using Schrödinger, free energy perturbation, and microsecond-scale MD simulations, enabling 100× screening efficiency, evaluating 1M+ compounds, and delivering three novel STAT6 chemotypes with sub-120 nM predicted affinity and favorable MM/GBSA scores, advancing two candidates to hit-to-lead stage.

Comparative Genomic Analysis of Beetle and Nematode Species

- Led de novo genome assembly and comparative genomics using PacBio HiFi, Hi-C, YAHs, BUSCO, and synteny analysis (RIdeogram, Jupiter plots), achieving 50 Mb N50 and 95% completeness, identifying species-specific genomic regions and conserved loci, and building a reproducible Nextflow pipeline for assembly validation and annotation (BRAKER, InterProScan).

SKILLS

Bioinformatics & Computational Biology : Spatial transcriptomics (NanoString CosMx), scRNA-seq analysis, Seurat, cell-state reference mapping, pseudotime and trajectory inference, differential expression analysis, multi-omics integration (scRNA-seq, bulk RNA-seq, ATAC-seq, Hi-C, ChIP-seq, CyTOF)

Computational Modeling & Systems Biology : Constraint-based metabolic modeling, COBRApy, flux balance analysis (FBA), metabolic-regulatory network integration, kinetic and perturbation simulations, sensitivity analysis, temporal graph modeling, CTMP models, protein–protein interaction (PPI) networks, link prediction

Molecular Modeling & Drug Discovery : Schrödinger suite, molecular docking, molecular dynamics simulations, free energy perturbation (FEP), MM/GBSA scoring, virtual screening, hit-to-lead optimization

Genomics & Comparative Genomics : PacBio HiFi assembly, Hi-C scaffolding, YAHs, purge_dups, Inspector, BUSCO evaluation, N50 optimization, repeat resolution, synteny analysis, variant calling, BRAKER annotation, InterProScan

Programming, Workflows & Infrastructure : Python, Bash, R (working knowledge), Linux, Docker, Nextflow, reproducible pipeline development, HPC computing, SLURM workload management, large-scale parallel simulations, automation and QC frameworks

Statistical & Machine Learning Methods : Gaussian Mixture Models (GMMs), logistic regression and calibration, feature engineering (ΔCq, ΔΔCq), model benchmarking, robustness testing, assay quality control and validation

Translational & Cross-Functional Skills : Translational bioinformatics, oncology and immunology research, clinical assay support, wet-lab collaboration, experimental design guidance, hypothesis generation, cross-functional stakeholder communication