Mahima Mahabaleshwar Siddheshwar

Bioinformatics Scientist | Genomics & Data Science | Computational Biologist

USA | (317) 506-5008 | mahima.siddheshwar@gmail.com | LinkedIn | Portfolio | https://github.com/MahimaMSiddheshwar

PROFESSIONAL SUMMARY

Bioinformatics Scientist with 3+ years of experience spearheading Next-Generation Sequencing (NGS) analysis, single-cell (scRNA-seq) and bulk RNA-seq to drive biomarker discovery and target identification. Skilled in building reproducible, scalable omics workflows in Python, R, and Bash, with a conceptual understanding of Docker and Nextflow/Snakemake for pipeline containerization. Experienced in leveraging AI/ML agents and statistical modeling to optimize biological reasoning across HPC and cloud environments. Includes hands-on wet-lab exposure to RNA extraction and NGS library prep to support R&D.

TECHNICAL SKILLS

- **Genomic Analysis & Pipelines:** scRNA-seq (Seurat, Scanpy), Bulk RNA-Seq (DESeq2, edgeR, limma), Multi-omics Integration, WGCNA, Variant Calling (GATK), Pathway Enrichment (GSEA, KEGG, Reactome), Spatial Transcriptomics and Omics Data.
- **Bioinformatics Tools & Resources:** STAR, BWA, SAMtools, BEDTools, ANNOVAR, Ensembl, GEO, TCGA, UCSC Genome Browser, Cytoscape, IGV
- **Programming & Data Science:** Python, R, Bash/Shell Scripting, SQL (Basic), Git Version Control
- **Machine Learning & Statistics:** PCA, t-SNE, UMAP, K-means, Hierarchical Clustering, LASSO Regression, SVM, Model Evaluation (ROC/AUC)
- **Cloud, HPC & Infrastructure:** HPC Cluster Computing (SLURM), AWS, Google Cloud Platform (GCP), Unix/Linux Workflow Automation
- **Data Visualization & Reporting:** Volcano Plots, Heatmaps, Interactive Dashboards (Plotly, Power BI, Tableau)
- **Wet Lab & Quality Control:** Cell Culture, DNA/RNA extraction, PCR, SDS-PAGE, FACS-based workflows, and NGS library preparation
- **Clinical, Quality & Regulatory:** HIPAA, GxP/GLP, ALCOA+, FAIR Principles, LIMS, FDA Documentation & Audit Readiness

PROFESSIONAL EXPERIENCE

Bioinformatics Analyst Intern – R&D | Elucidata Corporation, USA *July 2025 – October 2025*

- Optimized AI-assisted NGS workflows by ~30% via prompt refinement and LLM evaluation, improving biological reasoning for DEG and pathway analysis.
- Evaluated genomic workflow frameworks across 100+ curated research scenarios, ensuring biologically consistent and reproducible outputs before release.
- Performed LLM evaluation and prompt engineering, strengthening biological reasoning for KG, bulk/scRNA-seq, Synthea, and DEG tasks by 25–40%.
- Verified Synthea (synthetic health data) pipelines on cloud compute by confirming disease-specific cohort accuracy across 50+ repeatable test runs.
- Strengthened end-user research usability by stress-testing AI/ML agents across 100+ biological queries, improving genomics output quality & KG-driven insights.

Research Assistant – Global Health Data Analytics | IU BioHealth Informatics Research Center, USA *September 2023 – May 2025*

- Integrated team efforts to optimize RNA-seq preprocessing on HPC SLURM, cutting runtime by 40% using a Snakemake/Nextflow ready framework.
- Engineered reproducible RNA-seq pipelines using STAR, SAMtools, and BEDTools, generating high-quality, fully traceable gene-level feature matrices.
- Led QC & preprocessing automation using Bash scripting and SLURM batch jobs, improving pipeline stability by 25% & reducing manual team effort.
- Created pathway and network visualizations in Cytoscape, translating enrichment results into interpretable, biologically meaningful findings.

Bioinformatics Specialist | Maxgen Technologies (Client: Hetero Pharma), India *July 2021 – August 2023*

- Supported scRNA-seq clustering, annotation & biomarker analysis, improving insights by 25% and contributing to assay and preclinical decisions.
- Executed SOP approved RNA-seq & variant workflows under GxP/GLP/ALCOA+, ensuring 100% reproducibility & reducing gaps by 40%.
- Collaborated with R&D, engineering, and IT data teams to resolve pipeline issues, boosting workflow scalability by 20% and accelerating review cycles.
- Work hands-on with wet-lab scientists on PCR-based RNA extraction, NGS library preparation, and assay interpretation, delivering analyses that align with experimental goals and enable timely decision-making.

EDUCATION

Master of Science in Bioinformatics	August 2023 - May 2025
Luddy School of Informatics, Indiana University Indianapolis	Indiana, USA
Bachelor of Technology in Biotechnology	August 2016 - August 2020
Sapthagiri College of Engineering (Affiliated to VTU)	Bangalore, India

ACADEMIC PROJECTS

Machine Learning Pipeline for Breast Cancer Biomarker Discovery | Python (Pandas, NumPy, scikit-learn)

- Built a scalable ML framework to harmonize 7 GEO cohorts (~814 samples) using SVA/Limma, reducing batch-driven technical noise.
- Benchmarked five ML models (SVM, RF, ANN, LR, GBM) with stratified 80/20 validation, achieving 98.9% accuracy and AUROC >0.99.
- Validated a 28-gene signature on external datasets, identifying high-confidence biomarkers (POSTN, LSR) with $p < 1e-31$ for qPCR validation.

WGCNA on COVID-19 and RSV Transcriptomes | FastQC, HISAT2, SAMtools, R (Bioconductor, Tidyverse)

- Implemented an end-to-end RNA-seq pipeline (FastQC, HISAT2, featureCounts) on HPC, achieving >89% alignment across 68 samples.
- Applied WGCNA to identify key gene modules, showing strong correlation between MEyellow and COVID-19 severity ($r = 0.89$).
- Identified three shared antiviral hub genes (OASL, TXN, RBCK1) across COVID-19 and RSV datasets for downstream functional validation.

Streamlining RNA-Seq Data Analysis Framework | Linux shell scripting (Bash), HPC SLURM, Hierarchical Clustering, DEG

- Architected an automated RNA-seq processing & visualization framework on HPC, ensuring data integrity across 68 samples with >89% alignment rate.
- Profiled four pipelines for efficiency and quality, identifying t-SNE + Hierarchical Clustering as optimal (ARI = 0.60) with biological label agreement.
- Reduced computational runtime by 35% and applied the optimized framework to DEA, identifying 14,667 significant COVID-19 genes ($p < 0.05$).

PROFESSIONAL DEVELOPMENT & CERTIFICATIONS

- **Biosciences Program Certification — Biocon Academy × KGI (2021)**
  Industry training in biologics R&D, manufacturing, QC/QA, RA with hands-on lab, fermentation, mammalian cell culture, and drug-development projects.