# Mahima Mahabaleshwar Siddheshwar

**Bioinformatics Analyst | Genomics & Data Science**

USA | (317) 506-5008 | mahima.siddheshwar@gmail.com | LinkedIn | GitHub

## PROFESSIONAL SUMMARY

Bioinformatics professional with 3+ years of experience advancing R&D through AI/ML-integrated genomic workflows, RNA-Seq and scRNA-Seq data analysis, and LLM-driven pipelines for biomarker discovery and therapeutic insights. Expert in building and maintaining OMICS data-analysis pipelines using Python and R, with proficiency in statistical modeling, front-end tool development, and high-performance computing environments.

## TECHNICAL SKILLS

- **Bioinformatics & Genomics**: RNA-Seq/scRNA-Seq Pipelines (DESeq2, Seurat), WGCNA, GSEA, Feature Extraction, Genomic Data Integration (GEO, TCGA), OMICS, RNA-seq, scRNA-seq
- **Data Analysis & Programming**: Python (Pandas, NumPy, Scikit-learn, LLMs/Prompt Engineering), R (limma, sva, FastQC/TrimGalore/HISAT2), SQL, JMP, Power BI, Cytoscape; Statistical Modeling (LASSO, SVM, Linear Regression, Multivariate Analysis)
- **Computing & Visualizing Tools**: Unix/HPC (SLURM Clustering), Git for Version Control, t-SNE/PCA/K-means/Hierarchical Clustering, ROC/AUC/Volcano Plots/Heatmaps
- **Soft Skills**: Cross-Functional Collaboration, Technical Support & Training, Scientific Presentations, Troubleshooting & Optimization, Continuous Improvement

## PROFESSIONAL EXPERIENCE

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

- Architected LLM-powered RNA-Seq and genomics pipelines using Python/R, extracting features from 50+ datasets (including GEO/TCGA integrations) & updating databases with 100% traceability, reducing data inconsistencies by 25% and accelerating variant annotation/interpretation by 30%.
- Conducted integrative statistical analyses on multi-omics and clinical data, generating reproducible reports that informed 5+ R&D decisions, shortening analysis timelines by 35% via optimized workflows and multivariate modeling.
- Collaborated with 10+ scientists/engineers to evaluate and refine bioinformatics tools, achieving 20% accuracy gains through prompt engineering and scalability testing in high-performance computing environments.
- Supported data analysis for health trends and disease patterns, contributing to 3+ study designs with HIPAA/GDPR-compliant metadata management and ethical standards.
- Prepared interactive visualizations and summaries for stakeholders using Power BI and Cytoscape, providing on-demand technical support and training to ensure tool adoption and data quality.

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

- Integrated 100+ records from CDC/WHO repositories into standardized datasets, applying cross-referencing and statistical methods to eliminate 35% inconsistencies, enabling robust public health trend analysis and 20% faster cohort studies.
- Developed Power BI dashboards for geographic/disease burden monitoring from genomic data, visualizing multivariate indicators that reduced manual analysis time by 40% and supported 15+ R&D presentations.
- Documented and version-controlled (Git) bioinformatics pipelines for team collaboration, achieving 100% reproducibility in 20+ RNA-Seq analyses while upholding HIPAA/GDPR standards for data processing and interpretation.
- Collaborated with multidisciplinary teams, contributing to 5+ research protocols focused on health outcomes, global trends, and experiment design.

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

- Analyzed multi-omics datasets with Python/R/SQL/JMP/Power BI, applying LASSO/SVM modeling and Cytoscape network visualizations to identify trends, supporting R&D decisions and process enhancements by 15% in pharmaceutical workflows.
- Developed and maintained custom genomic data pipelines for biomarker discovery, integrating RNA-Seq features to streamline therapeutic insights and ensure reproducibility across 100+ analyses.
- Collaborated across R&D and data teams to troubleshoot pipeline performance, resolving issues that improved scalability by 20% and accelerated approvals in pharma environments.

## 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, R, LASSO/SVM | 2024**

- Integrated 7 GEO datasets (~350 samples) using limma/sva for normalization, reducing inconsistencies by 28%; applied LASSO for feature selection in multivariate analysis, achieving 93.4% accuracy in SVM/Logistic models with ROC/AUC validation.
- Identified 12 biomarkers via stratified cross-validation, enhancing generalizability by 32% and supporting RNA-Seq experiment design for clinical translation in precision medicine.

**WGCNA on COVID-19 and RSV Transcriptomes | RNA-Seq, R, WGCNA | 2024**

- Processed 68 RNA-Seq samples with FastQC/TrimGalore/HISAT2 pipelines on Unix/HPC, constructing WGCNA networks that identified 2 modules ($R^2$=0.89) correlating 90% with disease severity.
- Categorized 3 hub genes (OASL, TXN, RBCK1) with 95% confidence using GSEA; generated heatmaps/volcano plots/Venn diagrams, improving visualization clarity by 30% and reproducibility by 40%.

**Streamlining RNA-Seq Data Analysis Framework | Unix/HPC, SLURM, Clustering | 2024**

- Compared 4 pipelines (PCA/t-SNE/K-means/Hierarchical) on high-performance computing with SLURM, selecting t-SNE+Hierarchical for 40% better stability and reducing processing time by 35%.
- Containerized workflows for scalability, executing unit/integration tests to ensure quality in cloud-like environments, aligning with pharma-grade reproducibility standards.

## CERTIFICATIONS

- **Biosciences Program Certification** | Biocon Academy & Keck Graduate Institute (2021)
    - Experiential Learning in R&D, QC/QA, Production, Microbiology, Regulatory Affairs; Case Study on Drug Development; Hands-on Fermentation at BiOZEEN – Achieved 100% workflow mastery in bioscience operations.