

# Mahima Mahabaleshwar Siddheshwar

Indianapolis, IN | 317-506-5008 | [mahimamahabaleshwar@gmail.com](mailto:mahimamahabaleshwar@gmail.com) | <http://www.linkedin.com/in/mahima-m-siddheshwar-98a585196> | [GitHub](#)

## SUMMARY

Bioinformatics Scientist with a background in transcriptomics and multi-omic data curation and analysis, specializing in bulk and single-cell RNA-seq. Experienced in curating, checking, evaluating experimental design and processing transcriptomic datasets, hands-on skills in differential gene expression analysis, cell type annotation, gene set enrichment analysis, and results visualization. Demonstrated ability to develop and execute reproducible bioinformatics workflows in Unix/Linux and HPC environments. Proficient in R and Python programming, with a solid foundation in statistical concepts and NGS data integration. Adept at collaborating in cross-functional research teams and contributing to target identification and evaluation efforts.

## WORK EXPERIENCE

### Health Informatics Research Assistant

Bio Health Informatics Research Centre Indiana University

August 2023 – Present

- Streamlined organization and standardization processes for more than 30 distinct biomedical datasets, resulting in progressed data processing efficiency that contributed directly to faster turnarounds on critical research initiatives.
- Created interactive Power BI dashboards, reducing report preparation time by 40% and enhancing research decision-making efficiency.
- Utilized SQL-based data workflows, increasing database accessibility and consistency while reducing data processing errors by 25%.

### Quality Control (QC) Analyst

Biocon Biologics Limited

July 2021 – August 2023

- Reduced oncology drug stability analysis time by 20% through optimized data workflows, ensuring 100% compliance with GMP standards.
- Automated data integrity checks in LIMS and SAP systems, improving reporting accuracy by 15%.
- Led coordination of 15+ QC and R&D projects, increasing operational efficiency by 30% through streamlined communication, workflow optimization and achieving project milestones ahead of deadlines.
- Trained 10+ new joiners as a Subject Matter Expert (SME), improving onboarding efficiency and team productivity.
- Analyzed over 500+ in-process and stability samples using pre-established data analysis protocols, identifying trends and reducing errors by 15%.
- Improved incident reporting workflows, achieving a 30% boost in operational efficiency and strengthening adherence to 5S practices.

## SKILLS

**Bioinformatics Tools:** RNAseq, scRNAseq analysis, WGCNA, Next Generation Sequencing (NGS) pipelines (HISAT2, STAR, GATK, BWA), BLAST, Samtools, DESeq2, edgeR, GEO, TCGA

**Programming:** Python (Pandas, NumPy, scikit-learn, Matplotlib), R (ggplot2, dplyr, pheatmap), SQL, Bash Scripting, Machine Learning, Seurat, Scanpy

**Data Visualization:** Power BI, Tableau, Matplotlib, Seaborn, Plotly

**Computing and OS:** Unix/Linux, High-Performance Computing (HPC), SLURM, Jupyter Notebooks

**Data Handling:** GitHub, Workflow Automation

## EDUCATION

### Master of Science in Bioinformatics

Luddy School of Informatics, Indiana University Indianapolis

August 2023 - May 2025

Indiana, USA

### Bachelor of Technology in Biotechnology

Sapthagiri College of Engineering (Affiliated to VTU)

August 2016 - August 2020

Bangalore, India

## PROJECTS

### Comparative Transcriptomic Profiling of COVID-19 and RSV Using RNA-Seq and WGCNA

**Tools:** RNA-Seq Pipeline | WGCNA | HISAT2 | SAMtools | FastQC | FeatureCounts | R | Python | GEO Database | HPC

- Curated and processed 68 RNA-Seq samples (32 COVID-19, 34 healthy, 2 convalescent), reducing preprocessing time by 40% using FastQC, TrimGalore, HISAT2, and FeatureCounts.
- Constructed gene co-expression networks (WGCNA) and identified 2 key modules (MEyellow & MEgrey) with an  $R^2$  of 0.89, correlating 90% with COVID-19 severity.
- Identified 3 hub genes (OASL, TXN, RBCK1) with a biomarker confidence score of >95%, providing a 20% improvement in disease classification accuracy.
- Generated heatmaps, volcano plots, and Venn diagrams, enhancing data visualization clarity by 30% and improving biomedical research reproducibility by 40%.

### Streamlining RNA-Seq Data Analysis: A Framework for Comparing Computational Pipelines

**Tools:** RNA-Seq Pipeline | HPC | PCA | t-SNE | K-means | Hierarchical Clustering | Python | R | GEO Database

- Developed and compared four RNA-Seq clustering pipelines, integrating PCA, t-SNE, K-means, and Hierarchical Clustering, reducing data processing time by 30%.
- Executed pipelines in a Unix/Linux-based HPC environment, using SLURM for job scheduling reducing time by 35%.
- Determined t-SNE + Hierarchical Clustering as the most effective pipeline, achieving 40% better cluster stability and 30% improved biological relevance over conventional methods.