

Sarvesh Sunil Kurup

Boston | (857)351-3891 | kurup.sa@northeastern.edu | [LinkedIn](#) | [Github](#)

PROFILE

Bioinformatics master's student experienced in analyzing bulk and single-cell RNA-seq datasets using Python, R, and SLURM-based HPC clusters. Skilled in data integration, QC, statistical and machine learning methods for omics analysis and visualization. Strong track record of translating complex genomics data into biologically actionable insights through reproducible computational workflows.

EDUCATION

Northeastern University

Jan 2025 - Dec 2027

Master of Science, Bioinformatics (GPA: 4.0)

- **Coursework:** Bioinformatics Programming, Computational Methods in Bioinformatics, Omics in Bioinformatics, Biostatistics, Genomics, Data mining and Machine learning

Mumbai University

Sep 2020 - May 2024

Bachelor of Pharmacy, Pharmacy (GPA: 3.5)

- **Coursework:** Pharmaceutical Science, Biopharmaceutics, Medicinal Chemistry, Pharmaceutical Biotechnology

SKILLS

- **Computational & Programming:** Python, R, Bash, SQL, HPC (SLURM), Docker, Git, GitHub, Unix-like Environment
- **Molecular Biology & Omics Analysis:** Bulk RNA-seq, Single-cell RNA-seq(QC, normalization, clustering, batch correction, marker discovery), Differential expression, Pathway & gene-set analysis, Molecular Biology, Large-scale Omics Data
- **Software & Tools:** FastQC, HISAT2, Kallisto, Bowtie, Samtools, IGV, Scanpy, Seurat, Biopython, Streamlit, R Shiny
- **Machine Learning & Statistical Methods:** PCA, Clustering, Regression, Random forest, Bayesian modeling, HMMs, AI/ML Models

PROFESSIONAL EXPERIENCE

Northeastern University | *Teaching Assistant- Bioinformatics Programming(BINF 6200)*

Sep 2025 - May 2027

- Supported analysis and interpretation of large-scale omics datasets such as single-cell and bulk RNA-seq on HPC clusters using SLURM, emphasizing reproducibility and scientific rigor.
- Guided students in applying Python and R programming for bioinformatics tasks such as sequence analysis, data handling, and algorithm implementation.
- Provided one-on-one and group support, improving students' problem-solving, coding efficiency, and application of computational biology and molecular biology concepts.
- Assisted students with data integration, QC, and visualization workflows using Python and R within a unix-like environment for large biological datasets, leveraging modern data visualization tools.
- Translated complex computational concepts into accessible explanations for non-technical learners.

Sandoz(A Novartis Division) | *Intern- Drug Development and Quality Control*

May 2023 - Jun 2023

- Supported drug development analytics by integrating and analyzing pre-clinical molecular biology datasets using Python and R programming.
- Performed data QC, exploratory analysis, and interactive visualization using web-based tools to inform compound efficacy and safety assessment.
- Collaborated with multidisciplinary scientific teams to translate computational findings into biologically actionable insights, demonstrating strong communication and interpersonal skills.

PROJECTS

Bulk RNA-seq Analysis Pipeline (Bioconductor, KEGG) | [Github](#)

Jun 2025 - Jul 2025

- Analyzed RNA-seq datasets using R/Bioconductor to identify differentially expressed genes and dysregulated oncogenic and immune pathways, supporting target discovery and biological hypothesis generation.
- Optimized and scaled end-to-end RNA-seq workflows on HPC infrastructure for preclinical datasets, improving efficiency and reproducibility.
- Produced publication-ready visualizations to clearly communicate biological insights to cross-functional, non-computational collaborators.

Web-app for sc-RNA seq analysis (ScanPy, Streamlit) | [Github](#)

May 2025 - Jun 2025

- Developed a web-based, interactive single-cell RNA-seq visualization platform featuring UMAP/PCA embeddings, enabling intuitive exploration of cellular heterogeneity by non-computational users.
- Automated end-to-end scRNA-seq workflows, including quality control, clustering, and marker gene identification, substantially reducing analysis turnaround time.
- Integrated gene- and cell-type-specific expression visualizations to support interpretation of tumor heterogeneity and translational oncology research.

Bayesian Hierarchical Modeling for Clinical Trial Endpoints

Dec 2025

- Implemented Bayesian hierarchical models to estimate treatment effects across simulated multi-center clinical trials.
- Explored estimands, prior sensitivity, and posterior uncertainty to support regulatory-aware inference.
- Communicated findings through clear statistical narratives and visual summaries for non-technical audiences.
- Applied Bayesian modeling concepts relevant to early-stage therapeutic evaluation and translational decision-making.

ADDITIONAL EXPERIENCE & INTERESTS

- Exploring applications of LLM-based tools for automating data exploration, annotation, and interpretation of biological datasets.
- Strong interest in applying AI/ML and computational genomics to understand disease mechanisms and support drug discovery.