RISHABH TRIPATHI

Boston, MA — +1 617-217-8673 tripathirishabh91@gmail.com www.linkedin.com/in/rishabh-tripathi

Summary

Bioinformatics professional with a strong foundation in computational biology, genomics, and machine learning, focused on building end-to-end NGS analysis pipelines and decision-support tools for translational research. Experienced with RNA-seq, small RNA-seq, single-cell, spatial transcriptomics, and molecular docking, from raw data processing through statistical modeling, visualization, and interpretation. Comfortable in cross-functional, cloud-based environments using Python, R, Bash, Nextflow, Docker, Git/GitHub, and AWS to deliver reproducible, production-ready bioinformatics workflows.

Education

Northeastern University **Jan 2024 – May 2026**

Master of Science in Bioinformatics, GPA: 3.6 *Boston, MA*

Relevant coursework:Statistics for Bioinformatics, Genomics in Bioinformatics, Collect, Store, Retrieve, Analyze, Data, Omics, Transcriptomics.

Loyola Academy Degree and PG College **Aug 2020 – May 2023**

Bachelor of Science in Biotechnology, Chemistry, and Genetics, GPA: 3.1 *Hyderabad, India*

Relevant coursework: Plant Biotechnology, Molecular Biology, Agriculture Biotechnology, Immunology, Bioanalytical Techniques

Experience

Bioinformatics CO-OP **July 2025 – Dec 2025**

Camp4 Therapeutics Corporation *Boston, MA*

- Designed and maintained reproducible NGS bioinformatics workflows in R, Python, and Bash, building custom data processing, QC, and visualization scripts and **deploying R Shiny web apps** for exploratory and summary-level analysis.
- Processed bulk **RNA-seq and ChIP-seq** datasets with production-grade Nextflow NGS pipelines on **AWS EC2**, supporting **time-course transcriptomics**, functional genomics, and gene regulatory analyses through parameterized, scalable workflow orchestration.
- Built and enhanced internal analytics and visualization tools for **target discovery** and **ASO screening**, including interactive dashboards, EC50 **dose–response analyses**, **sequence-alignment visualizations**, and filterable UI, collaborating cross-functionally with scientists, data engineers, and chemistry team.

RNA Sequencing Pipeline Developer **Jan 2024 – May 2024**

Northeastern University *Boston, MA*

- Designed and optimized **a containerized small RNA-seq pipeline** modeled after *miRge3.0*, integrating **UMI-aware read collapsing**, **isomiR profiling**, and **parallelized alignment** to enhance analytical accuracy and achieve a **30% reduction in computational runtime**.
- Automated **NGS data processing** (QC, adapter trimming, alignment) using **Python**, **Bash**, and **Docker**, with continuous integration via **GitHub Actions** to ensure **reproducibility**, **traceability**, and **pipeline robustness**.
- Implemented **cloud-native sequencing workflows** leveraging **AWS S3** and **AWS Batch** for scalable compute and built R Shiny + Highcharts applications for **interactive visualization of miRNA/isomiR distributions and QC summaries**.

Undergraduate Research Intern – Molecular Docking **Nov 2022 – Mar 2023**

Ciencia Life Sciences *Hyderabad, India*

- Performed **structure-based virtual screening and molecular docking** targeting **TMPRSS4 (Homo sapiens)** using **AutoDock Vina** and **iGEMDOCK**, followed by **force-field-based energy minimization** and **binding affinity scoring** to prioritize candidate inhibitors.
- Curated and integrated **natural product and small-molecule libraries** from **ZINC** and **COCONUT databases**, applying **drug-likeness, ADME, and toxicity filtering** via **SwissADME**, reducing ~10,000 compounds to **~5,000 drug-like ligands** for downstream docking.
- Conducted **ligand preparation**, **homology modeling**, **and structural validation** (Ramachandran analysis), and analyzed **protein-ligand, interactions**, **binding conformations, and active-site residues** using **UCSF Chimera** and **BIOVIA Discovery Studio**.

Projects

Cell-Type–Resolved Spatial Transcriptomics Atlas (Mouse Brain, Visium HD) **Dec 2025**

Spatial transcriptomics • scRNA-seq integration • Deconvolution • Domain analysis GitHub

- Built an **end-to-end spatial transcriptomics atlas** by integrating **Visium HD (8 µm & 16 µm) spatial RNA-seq data** with **single-cell RNA-seq references**, performing **cell-type deconvolution and label transfer** to generate **per-bin cell-type abundance maps** across tissue sections.
- Developed a spatial–expression domain pipeline with graph-based smoothing and Leiden clustering to detect coherent spatial domains and domain-specific marker genes via non-parametric differential expression.
- Implemented QC and enrichment analytics (UMI depth, gene complexity, mitochondrial content, domain-level enrichment) and exported standardized AnnData/Parquet/CSV outputs, interactive dashboards, and per-sample QC HTML for reproducible atlas generation

Interpretable Drug Response Prediction (DepMap / PRISM) **Aug 2025**

Interpretable Machine Learning • Genomics • Drug Response Modeling GitHub

- Built a leakage-safe ML pipeline to predict drug response from bulk RNA-seq gene expression (DepMap/CCLE) and PRISM primary screen data using Elastic Net and XGBoost.
- Applied **SHAP-based interpretability** and **pathway aggregation (GO/Reactome/KEGG)** to convert model predictions into mechanistic biological Insights, identifying key cholesterol and lipid metabolism drivers.
- Demonstrated **class-consistent biological explanations across four statins**, identifying shared driver genes (**SQLE, HMGCS1, GGPS1**) and conserved cholesterol-related pathways.

Single-Cell RNA-Seq Analysis & Cell Type Classification Pipeline **March 2025**

Python, Scanpy, scikit-learn, TensorFlow, GSEApy GitHub

- Built **scRNA-seq pipeline processing 10x Genomics PBMC** data through **QC, normalization, clustering and UMAP/t-SNE** using **Python** and **Scanpy;** engineered **cell type classification using Random Forest, SVM, and TensorFlow with SMOTE balancing**; implemented **GO and KEGG** pathway **enrichment** with GSEApy generating cross-cluster heatmaps and marker gene identification; **developed modular CLI for reproducible preprocessing, differential expression, and ML training workflows**.
- Optimized pipeline runtime to process 4,000+ cells through complete analysis workflow in 8 minutes on standard hardware.