PRANAVA UPPARLAPALLI

Leesburg, VA | +1 (945) 527-5490 | pxu.bioinfo@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

Bioinformatics engineer with 2+ years building reproducible next-generation sequencing (NGS) and transcriptomics workflows for RNA-seq, TWAS, single-cell, and 10x Visium spatial data. Combines Python/R, Nextflow, Linux, and SLURM high-performance computing (HPC) to automate quality control (QC), troubleshoot pipeline failures, and deliver interpretable cohort-scale outputs for research and biotech teams.

EXPERIENCE

University of Texas at Dallas | *Bioinformatics Analyst* — Jan 2025 - May 2025

- Ran RNA-seq and TWAS pipelines across ~1,100 human brain cortex samples on Linux SLURM HPC, delivering cohort-level QC outputs and traceable run records for research collaborators.
- Resolved Nextflow failures involving memory limits, malformed inputs, and partial batches, improving rerun decisions and pipeline reliability across cohort stages.
- Built Python QC tooling with pandas/matplotlib to validate outputs, flag normalization issues, and standardize run-level quality checks.
- Presented QC findings, pipeline issues, and analysis outputs through weekly reports and monthly PPT meetings, supporting decisions across multi-investigator genomics projects.
- Documented pipeline parameters, run notes, and QC outputs to support reproducible re-execution across cohort stages.
- Used Git-versioned documentation and structured outputs to improve traceability across RNA-seq and TWAS workflow runs.

Bharati Vidyapeeth University | *Bioinformatics Research Associate* — Jun 2022 - May 2023

- Prepared bulk RNA-seq libraries and PCR-amplified genomic targets using Illumina TruSeq kits, enabling downstream NGS workflows and supporting computational RNA-seq analysis that identified key expression patterns for the project
- Analyzed bulk RNA-seq datasets in R, generating QC summaries, normalized expression outputs, differential expression results, and visualizations for research review.
- Built reusable R and Bash scripts for QC, normalization, differential expression, and GSEA on public cancer transcriptomics datasets, cutting analysis setup time by about 30 % and ensuring reproducible pipelines
- Delivered annotated R Markdown reports with volcano plots, PCA, pathway enrichment, and gene ontology interpretation for cohort review.

Sree Vidyanikethan Degree College | *Microbiologist* — Aug 2021 - Mar 2022

- Performed disk diffusion and MIC assays on 8 bacterial strains per CLSI SOPs, flagging 3 MDR isolates.
- Centralized SOPs and records across 4 projects, reducing protocol deviations by ~30% and improving reproducibility

PROJECTS

Spatial Transcriptomics Cohort Pipeline | GitHub: Spatial analysis — Feb 2026

- Built a version-controlled Nextflow/R pipeline for 10x Visium spatial transcriptomics, converting Visium outputs to SpatialExperiment objects and running spot QC, normalization, BayesSpace domain detection, Moran's I SVG ranking, and cohort summaries.
- Validated spatial domains with ARI/stability 0.71 and built Shiny/plotly outputs for multi-sample QC and spatial review.

Colorectal Cancer TME Single-Cell Analysis | GitHub: scAnalysis — Oct 2025

- Processed 63,689 scRNA-seq cells from GSE132465 through QC, scVI integration, Leiden clustering, UMAP visualization, and manual cell-type annotation.
- Identified 31 clusters and 8 major cell types, highlighting myeloid/stromal shifts and exhausted T-cell signatures consistent with an immunosuppressive colorectal TME.

nf-core/rnaseq Run-Risk Predictor | GitHub: run-risk — Jan 2025

- Built a Python CLI/package that parses nf-core/rnaseq Nextflow trace/log outputs into structured features, trains scikit-learn logistic regression models to flag run-failure and resource-risk patterns, and validates parsers/dataset builders with pytest.

TinyVariant | GitHub: TinyVariant — Sep 2025

- Built a ClinVar variant-classification proof of concept comparing scikit-learn logistic regression with a PyTorch Tiny Recursive Model on ~100k variants.
- Found logistic regression remained stronger than the TRM model (ROC AUC 0.959 vs 0.951), documenting ablations and model tradeoffs for reproducible evaluation.

SKILLS

- **Languages:** Python, R, Bash, Unix/Linux, Git, structured query language (SQL), PostgreSQL, Jupyter, RMarkdown, pandas
- **Workflow & Infra:** Nextflow, nf-core, Snakemake, reproducible workflows, Docker, Singularity, Conda, pipeline validation, software testing, pytest
- **HPC & Cloud:** SLURM HPC, AWS fundamentals (AWS 101), Linux systems, FAIR workflows
- **NGS & Omics:** RNA-seq, scRNA-seq, snRNA-seq, ATAC-seq, spatial transcriptomics (10x Visium), TWAS, multi-omics
- **Alignment & QC:** STAR, salmon, kallisto, FastQC, MultiQC, Samtools, Bedtools, IGV, PrediXcan
- **Omics Tools:** Scanpy, Seurat, scVI, SpatialExperiment, BayesSpace, Moran's I, Harmony, DESeq2, edgeR, limma, CellChat
- **Databases:** GEO, SRA, TCGA, NCBI, Ensembl, ENCODE, UCSC, KEGG, cBioPortal
- **Machine Learning (ML) & Stats:** scikit-learn, PyTorch, GSEA, pathway enrichment, GO, PCA, clustering, differential expression
- **Visualization:** matplotlib, seaborn, ggplot2, R Shiny, Plotly, UMAP/tSNE

EDUCATION

M.S., Bioinformatics and Computational Biology — **Aug 2023 - May 2025**
University of Texas at Dallas — *Richardson, TX*

B.S., Microbiology, Biochemistry & Chemistry — **Jun 2019 - May 2022**
Sri Venkateshwara University — *Tirupathi, India*