Shivansh Verma

Bioinformatician & Data Analyst
Bengaluru,Karnataka | P: 8800573508 | Email: shiv.v10899@gmail.com
LinkedIn: https://www.linkedin.com/in/shivansh-verma05/ | Github: https://github.com/shivanshv9911
Portfolio: https://shivanshv9911.github.io/shivansh-verma-portfolio/

PROFESSIONAL SUMMARY

Bioinformatics Data Analyst with 2+ years of experience building scalable NGS analysis pipelines and dashboards to support experimental biology teams. Proficient in Python, Snakemake, DESeq2, Google Cloud, and data visualization tools. Known for automating time-intensive workflows and integrating qPCR, RNA-seq, HybELISA, and Cell Autotolerability assay data to generate actionable biological insights. Adept at collaborating with global teams to accelerate research timelines in drug discovery.

EDUCATION

Indraprastha Institute of Information Technology Delhi (IIITD) New Delhi,
M.Tech – Computational Biology Aug 2021-Aug 2023
CGPA : 9.0/10
Thesis Title : Analysis of Gene Expression and Structural Variations in Common Genes Across Populations: Implications for Cardiovascular Disease and Reverse Cholesterol Pathway

Noida Institute of Engineering and Technology (NIET) Greater Noida,UP
B.Tech – Biotechnology Jul 2017 - Jul 2021
CGPA : 8.9/10

PROFESSIONAL WORK EXPERIENCE

Creyon Bio, India (Startup)
Data Analyst / Research Associate – Data Analysis [August 2023 – Present | Full-time | Bengaluru, India]

1. **Automated NGS Analysis Pipelines (Differential Gene Expression (DGE), Nanopore Pipeline)**
   - Built and deployed modular pipelines using **Python and Snakemake** for bulk RNA-seq and **Nanopore data**.
   - Integrated DESeq2-based workflows with downstream visualization in R/Python, improving reproducibility and **reducing turnaround by 40%.**
   - Automated Nanopore QC integrated with Google Cloud enabled scalable and reliable storage and access.
   - **Key Impact:** Delivered standardized pipelines adopted across internal experiments.
2. **Dashboard Development & Reporting Automation (qPCR, HybELISA)**
   - Designed reproducible workflows and developed **Streamlit dashboards** for real-time qPCR and HybELISA assay reporting.
   - Eliminated manual reporting, saving **10+ hours/week** and improving transparency for wet lab teams.
   - **Key Impact**: Enabled rapid, error-free tracking of assay metrics across timepoints.
3. **Local Infrastructure Setup & Team Enablement (QNAP Server)**
   - Supported the implementation of a **QNAP local storage system** alongside the vendor team; documented SOPs and gained hands-on experience in managing secure, on-premise data infrastructure for bioinformatics workflows.
   - Helped establish SOPs and trained cross-functional teams on structured data upload and access protocols.
   - **Key Impact:** Improved team-wide data consistency and enabled reliable local storage workflows.
4. **In Vivo Data Integration & Machine Learning Preparation**
   - Consolidated **in vivo datasets from 20+ sources** into a harmonized, ML-ready format with built-in QC.
   - Streamlined downstream modeling efforts through standardized data pipelines.
   - **Key Impact:** Enabled early-stage predictive models for assay outcome interpretation.

5. **Collaboration & Documentation for Assay Development**
    - Partnered with biologists and data scientists (India/US) to design assays and interpret integrated datasets.
    - Documented workflows and dashboards to support reproducibility, handoffs, and scaling.
    - **Key Impact:** Facilitated smoother collaboration and tech transfer across teams.

ACADEMIC & RESEARCH PROJECTS

1. **Comprehensive Analysis of RNA Sequencing, Connectivity Mapping, and Molecular Docking for Cardiovascular Disease Therapeutics**
   [Academic Project – IIIT Delhi | Jan 2023 – May 2023]
    - Analyzed RNA-seq data to identify differentially expressed genes and pathways.
    - Applied Connectivity Mapping to discover small molecules capable of reversing disease signatures.
    - Validated drug candidates through molecular docking simulations.
    - **Tools:** AutoDock, AutoDock Vina, DESeq2
    - **Key Metric:** Identified top 10 candidate molecules with strong binding affinities.
2. **Automated Web Scraping for Protein Dataset Development for Machine Learning Models**
   [Independent Project | Dec 2022]
    - Built a BeautifulSoup-based web scraper to extract structured protein metadata from the RCSB database.
    - Created a curated dataset ready for machine learning model training and evaluation.
    - **Tools:** Python (BeautifulSoup)
    - **Key Metric:** Extracted and cleaned 5,000+ protein entries.
3. **Differential Gene Expression and Pathway Enrichment Analysis in the Reverse Cholesterol Pathway**
   [Academic Project – IIIT Delhi | Aug 2022 – Dec 2022]
    - Developed Bash scripts and Python workflows for bulk RNA-seq preprocessing and differential expression analysis.
    - Conducted gene set enrichment analysis focusing on cholesterol metabolism pathways.
    - **Tools:** Bash, Python (Seaborn, Matplotlib), DESeq2
    - **Key Metric:** Identified 200+ differentially expressed genes associated with cholesterol metabolism.
4. **Classification of Protein Interaction Patterns Using Machine Learning**
   [Academic Project – IIIT Delhi | Aug 2022 – Dec 2022]
    - Developed machine learning models to classify interacting vs. non-interacting protein patterns.
    - Deployed the trained model into an interactive web application using Streamlit.
    - **Tools:** Python (Scikit-learn, Streamlit)
    - **Key Metric:** Achieved 85% classification accuracy on test datasets.

TECHNICAL SKILLS

- **Programming & Data Analysis:** Python, R, SQL, Pandas, NumPy, Jupyter Notebooks
- **Bioinformatics Tools & Pipelines:** Snakemake, Nextflow**,** DESeq2 (Python), FastQC, MultiQC, NanoPlot, STAR, HISAT2, featureCounts
- **Cloud & Infrastructure:** Google Cloud Platform (BigQuery), Docker, Git, **Linux**, Bash,**CI/CD pipelines**
- **Single-cell & Spatial Transcriptomics Analysis**: **Seurat (R), ScanPy (Python), Squidpy (Python), Voyage (R)**
- **Data Visualization & Dashboarding:** Looker, Streamlit, Seaborn, Matplotlib, Excel (Advanced)

Awards: 3rd Prize – Startup Conclave (IMS Ghaziabad, 2018); Department Topper Award (2020); TCS NQT Qualified (2020); GATE-2021 (Biotech).
Languages: Fluent in English and Hindi; Beginner in German
Certifications & Training: GCP Data Engineering (Coursera), Docker for the Absolute Beginner - Hands On - DevOps (Udemy), Data Science with Python (IBM)
Conferences: Attended illumina Genomics Summit, Bengaluru (30th August 2024)

ACTIVITIES

- **Teaching Assistant, IIITD** – Supported instruction and student learning.
- **Student Welfare Committee, NIET** – Editorial and event coordinator, launched departmental magazine and led 5+ annual events.