Nikhil Mark Lakra

 Noida, U.P.   nikhilmarklakra@gmail.com   +91-9599895356   in LinkedIn   ○ GitHub

Experience

Bioinformatics Scientist   *Delhi*
Elucidata   *Oct 2022 – Present*

- Collaborated with major pharmaceutical clients on diverse research projects, leveraging Bulk RNA-Seq and single-cell transcriptomics data to derive high-quality insights for early drug discovery.
- Standardized and automated workflows for preprocessing, differential expression analysis, and pathway enrichment, ensuring reproducible and efficient results delivery.
- Developed and deployed a meta-analysis pipeline with KNN-based gene imputation for differential expression analysis, and pathway enrichment using Python and R, enabling the discovery of key therapeutic targets.
- Built a Dockerized gene similarity pipeline to identify genes with similar expression signature patterns in single-cell RNA-seq datasets, optimized for datasets averaging 100k cells. Using tools such as Pandas, Scikit-learn, Scanpy, and scvi-tools to deliver an efficient, scalable, and reproducible analysis workflow and contributing to a 40% increase in the value of the renewal contract.
- Developed a workflow using pre-trained machine learning models (e.g., scVI-tools, CellTypist) to automate the annotation of cell types in scRNA-Seq data, reducing the time to identify and accurately annotate cell types by 20%.
- Conducted data audits on public repositories (e.g., GEO, Single Cell Portal, and ArrayExpress) to identify high-quality datasets of interest to support client-specific analyses and projects.
- Led an internal training session on preprocessing bulk RNA-seq data, covering the nuances of normalization methods and the underlying algorithms of differential expression tools (LIMMA, edgeR, DESeq2), significantly enhancing the team's proficiency in downstream analysis workflows.

Education

Indraprastha Institute of Information Technology Delhi (IIITD)   *Oct 2019 – Jan 2022*
M.Tech in Computational Biology

- GPA: 7.96/10.0
- **Coursework:** Machine Learning, Data Science in Genomics, Biostatistics, Computer Aided Drug Design

Projects and Thesis

Thesis: Predicting Selection Pressure on SNPs in Human Populations   *Link to Thesis*

- Developed a machine learning pipeline to predict single nucleotide polymorphisms (SNPs) under selection pressure across 17 populations from the 1000 Genomes Project.
- Implemented a Random Forest classifier, taking advantage of population genetic characteristics such as allele frequency, F statistics, and linkage disequilibrium for accurate predictions.
- Achieved high classification accuracy, contributing to the identification of genetic variations of evolutionary significance.

AQI Forecasting Using Deep Learning (LSTM)   *GitHub Link*

- Developed an LSTM-based deep learning model in PyTorch to predict the next day's Air Quality Index (AQI) using historical AQI data from CPCB and climate data from NASA Power.

Technologies

Languages: Python, R, SQL, Bash

Technologies & Frameworks: Pandas, NumPy, scikit-learn, TensorFlow, PyTorch, scVI-tools, Scanpy, Seurat, DESeq2, edgeR, LIMMA

Cloud & Development Tools: Docker, Git, AWS, Snakemake, PostgreSQL, SQLite