

MAYANK MURALI

(814)-777-7430

MAYANKMURALI12@GMAIL.COM

www.linkedin.com/in/mayankmurali

WORK EXPERIENCE

- | | | |
|---|---|------------------------------|
| Senior Data Analyst | The University of Virginia | July 2022 – Present |
| <ul style="list-style-type: none">Led data wrangling and pre-processing efforts for different research projects on sequencing datasets. Executed in-house Nextflow pipeline using Docker to derive protein isoform models from raw PacBio long-read RNA-Seq data (Ir-Seq).Wrote Python scripts for designing UCSC browser tracks, visualizing the effect of splicing QTLs on protein isoforms for coronary artery disease and osteoporosis. Collaborated with research labs in a cross-functional environment.Analyzed and quantified gene and isoform expression using PacBio Ir-Seq data. Familiar with using RNA-seq tools such as STAR, RSEM, Kallisto, LeafCutter, Mfuzz, and Jutils, as well as the IsoSeq workflow for MAS-Seq single-cell analysis.Performed isoform-informative peptide analysis using proteomics-based tools like MetaMorpheus and PoGo. | | |
| Graduate Teaching Assistant | The Pennsylvania State University | Jan 2020 – Dec 2021 |
| <ul style="list-style-type: none">Facilitated cognitive learning on 'Introduction to System Programming (CMPSC 311)' course for B.S. (Engineering) students for 20 credit hours which included tutoring sessions, grading as well as counselling students in need. | | |
| Full Stack Developer Intern | Agrometrics Analytics & Technology Pvt. Ltd. | Jan 2019 – April 2019 |
| <ul style="list-style-type: none">Developed backend software modules that capture large scale farm operations data from SQL database and relay them to the front-end website interface using Python, PHP and SQL as part a web application. | | |

PROJECTS

- Biosurfer** | [GitHub](#) | *Python*
A Python framework leveraging full-length protein isoform data from PacBio Ir-Seq to catalogue and visualize complex alternative splicing events, resulting in protein sequence variations. Utilized SQLite for efficient data management.
- M.S Thesis** | [GitHub](#)
Designed and implemented a pipeline for benchmarking and analyzing de novo metagenome assembly tools utilizing long-read PacBio HiFi datasets. Evaluated misassembly rate, quality, and completeness of metagenome assembled genomes. Developed bash scripts to execute assemblers, assess assembly outcomes, and visualize contigs through Bandage tool on a high-performance computing (HPC) cluster.
- A study on *k*-mer counting methods using different data structures** | *C++*
Evaluated the performance of unordered map, suffix array, hash table, and bloom filter in the context of KMC 2 algorithm for *k*-mer counting on human reference genome and a bacterium datasets. Benchmarking focused on memory usage and runtime.
- Mouse FOXC1 gene regulation analysis**
Executed a Galaxy pipeline for ChIP-exo and RNA-seq analysis. Investigated FOXC1 transcription factor binding sites in mouse stem cell data and explored FOXC1 gene regulation.

EDUCATION

- | | | |
|---|----------------------------|-----------------------------|
| The Pennsylvania State University | University Park, PA | Dec 2019 - June 2022 |
| <ul style="list-style-type: none"><i>Master of Science, Computer Science, GPA: 3.25/4.0</i><i>Selected Coursework:</i> Algorithm Design and Analysis, Fundamentals of Computer Architecture, Digital Image Processing II, Pattern Recognition and Machine Learning, Computational Biology, Algorithms and Data Structures in Bioinformatics. | | |
| Vellore Institute of Technology | Chennai, India | 2015 - 2019 |
| <i>Bachelor of Technology, Computer Science and Engineering, GPA: 8.37/10</i> | | |

LANGUAGES AND TECHNOLOGIES

- Programming/Scripting Languages:* C, C++, Python, R, Bash, Java, SQL, HTML, PHP, JavaScript/TypeScript.
- Framework and tools:* Git, Excel, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, MATLAB, LaTeX, Visual Studio, Jupyter, RStudio, Anaconda, Galaxy, Nextflow, Snakemake, Docker.

RESEARCH PUBLICATIONS

- Murali .M et al.(2022). Biosurfer: Connecting genomic, transcriptomic, and proteomic layers to characterize protein isoform function.** In draft.
- Abdullah A.,..., Murali .M et al.(2023). Long-read proteogenomics to connect disease-associated sQTLs to the protein isoform effectors of disease.** bioRxiv 2023.03.17.531557; doi: <https://doi.org/10.1101/2023.03.17.531557>
- Mehlferber M.M, Jeffery .E,., Murali .M, et al (2022). Characterization of protein isoform diversity in human umbilical vein endothelial cells via long-read proteogenomics.** *RNA Biology*, Vol 19. doi: 10.1080/15476286.2022.2141938
- M. Murali et al (2019). Data Analytics on IoT-based Health monitoring system.** *International Journal of Recent Technology & Engineering* (Scopus- indexed) – ISSN:2277-3878, Volume-8, Issue-1, pp. 220-223.
- B. Sahoo, A. Maharana, M. Murali et al. (2019). Low-Cost Air Sensing System.** In *Proceedings of 3rd International Conference on Computing and Communications Technologies (ICCCCT)*. IEEE. doi: 10.1109/ICCCCT2.2019.8824890.