

# MAYANK MURALI

(814)-777-7430

MAYANKMURALI12@GMAIL.COM

[www.linkedin.com/in/mayankmurali](https://www.linkedin.com/in/mayankmurali)

## WORK EXPERIENCE

### Computational Associate II

Broad Institute of MIT and Harvard

Nov 2023 – Present

- Developed pipelines for genotyping, calculating PRS and LDSC heritability on Terra platform for in-house biobank within the lab.
- Developed computational framework to identify effect of variants through PRS on TF motifs across disease risk in Adipose tissue using single-cell ATAC seq data.
- Performed population genetics-related analysis on FinnGen.

### Senior Data Analyst

The University of Virginia

July

2022 – Nov 2023

- Developed Python framework (Biosurfer) that uses full-length protein isoform knowledge from LRS data to catalogue (using SQLite) and visualize the complex alternative splicing events that lead to protein sequence changes.
- Involved in running the Nextflow pipeline with Docker to generate IsoSeq files.
- Developed Python scripts to design UCSC browser tracks that visualizes isoforms of having lead sQTLs in genes of interest based on abundance.

### Graduate Teaching Assistant

The Pennsylvania State University

Jan 2020 – Dec 2021

- 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

- 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.

## ACADEMIC PROJECTS

### M.S Thesis | [GitHub](#)

Developed pipeline to benchmark and analyse *de novo* metagenome assembly tools on the long-read PacBio HiFi datasets based on misassembly rate, quality and completeness of metagenome assembled genomes. Wrote bash scripts to run the assemblers, evaluate the assembly and visualize the contigs (using Bandage) on HPC cluster.

### A study on *k*-mer counting methods using different data structures | C++

Analysed the performance of unordered map, suffix array, hash table and bloom filter with respect to KMC 2 algorithm on *k*-mer counting operation using the human genome GRCh38 and the *Staphylococcus aureus* bacterium datasets. The benchmarking was performed based on the memory usage and runtime.

### Understanding FOXC1 mouse gene regulation

Implemented a pipeline on Galaxy to perform ChIP-exo and RNA-seq analysis. Identified the binding sites of the transcription factor protein Forkhead Box C1 (FOXC1) in mouse stem cell data and investigated the regulation of the FOXC1 gene.

## EDUCATION

The Pennsylvania State University

University Park, PA

Dec 2019 - June 2022

- Master of Science, Computer Science, GPA: 3.25/4.0
- Selected Coursework: 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

Bachelor of Technology, Computer Science and Engineering, GPA: 8.37/10

## LANGUAGES AND TECHNOLOGIES

- Programming/Scripting Languages: C, C++, Python, Bash, Java, SQL, HTML, PHP, JavaScript/TypeScript, R.
- Framework and tools: Git, MATLAB, LaTeX, Visual Studio, Jupyter, Anaconda, Galaxy, Nextflow, Docker, HPC.

## ADDITIONAL EXPERIENCE AND AWARDS

- TEDx Event Management:** Executive member of Organization Committee of TEDx VIT Chennai hosted by the English Literary Association.

- **ENACTUS for Social Uplift, Project Lead:** Student Member of Enactus VIT Chennai (2017-2018), took part in implementation of community outreach projects for economic upliftment of the underprivileged in India.

#### **RESEARCH PUBLICATIONS**

---

- **Murali M**, Saquing J, et al. (2025). Biosurfer: Connecting genomic, transcriptomic, and proteomic layers to characterize protein isoform function. *Genome Research*, 35, 1012-1024. doi:10.1101/gr.279317.124 | [GitHub](#)
- Lu. S.,.. **Murali .M** et al. (2024). Predicting physiological stability of alternatively spliced protein isoforms using integrated sequence-based feature representation. Under review for ISMB proceedings.
- Abood A, **Murali. M** et al. (2024). Long-read proteogenomics to connect disease-associated sQTLs to the protein isoform effectors of disease. *\*American Journal of Human Genetics\**, 111(9), 1914-1931. doi:10.1016/j.ajhg.2024.07.003
- 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 (ICCCT)*. IEEE. doi: 10.1109/ICCCT2.2019.8824890