REVATHY VENUKUTTAN

+1-312-613-2471 | revathyvenukuttan@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Bioinformatics professional with 3 years of experience in computational functional genomics, inferential statistics, and machine learning. Highly skilled in bioinformatic tools, computational software, and pipeline development in HPC and Cloud environments. Proficient in programming languages such as R, Bioconductor, Python for advanced data analysis. Adept at managing and executing projects of varying complexity, ensuring timely delivery and adherence to scientific best practices. Good communication and collaboration skills, enabling seamless interaction with cross-functional teams and stakeholders.

EDUCATIONAL QUALIFICATIONS

University of Illinois at Chicago (UIC), IL   *May 2020*
Master of Science in Bioinformatics (CGPA: 3.71/4.00)
Relevant Coursework: Computation Genomics, Statistics and Machine Learning, Advanced Statistics and Machine Learning, Cell Biology, Molecular Biology

College of Engineering, Trivandrum (CET) / University of Kerala, India   *Jun 2017*
Bachelor of Technology in Industrial Engineering (CGPA: 8.1/10.0)
Relevant Coursework: Industrial Statistics, Object Oriented Programming, System Analysis and Design, Stochastics

PROFESSIONAL EXPERIENCE

Bioinformatician I
Center for Advanced Genomic Technologies, Duke University   *Aug '20-May '24*

o Experience with multi-omics (WGS, RNA-seq, ATAC-seq, STARR-seq, ChIP-seq, scRNA-seq) data and perform large scale genomic analysis including QC, data preprocessing, downstream analysis, and result visualization.
o Perform integrative analysis on high-throughput transcriptomic and epigenomic data to identify potential guide RNA targets for CRISPRi/a screens in mouse immune cells to study the activation and differentiation profiles of T cells.
o Implement a joint analysis model that integrates genetic, epigenomic, and phenotypic data from public databases (GWAS, ENCODE, 1000Genomes) to build a reporter assay library, identifying genetic mediators of auto-immune disorders.
o Develop and maintain preprocessing and analysis pipelines in CWL by staying up to date with algorithms, data formats and best practices.
o Collaborate with researchers to write custom functional scripts in R and Python upstream of analysis workflows as well as repackage and document existing scripts for reproducibility and ease of maintenance.
o Active involvement in working groups and jamborees to standardize data formats, processing techniques and result comparisons for MPRA, Single-Cell and CRISPR data as part of ENCODE and IGVF consortiums.
o Submit and release experimental data to public databases like ENCODE, IGVF and GEO for genomic assays (RNA-seq) and functional characterization assays (ATAC-seq, STARR-seq).
o Continuously contribute to paper presentations, poster presentations, research grants and manuscripts under the supervision of mentors and peers.
o Involve in training researchers across the center in Unix & HPC scripting, bioinformatics tools and hands-on sessions for running workflows and pipelines.
o Co-organize the design of training workshops that introduces genetics, genomics, and bioinformatics concepts to participants as part of outreach activities of the center.

Research Assistant
Department of Biological Sciences, UIC   *Feb '20-Aug '20*

o Extracted and categorized FASTA sequences of moonlighting proteins based on organism, gram stain from MoonProt - a database for Moonlighting proteins.
o Performed motif discovery and enrichment analysis using MEME suite to find similarities between categories of moonlighting proteins.

Research Assistant
Department of Neurology and Rehabilitation, UIC   *Feb '19-Feb '20*
   o   Developed semi-supervised algorithm in R for feature extraction and background correction for microarray data.
   o   Assessed gene expression variations and identified differentially expressed genes (limma-voom) between samples from different regions of spinal cord for human ALS patient samples.

Graduate Research Assistant–Data Sciences
OSF-UIC Research Lab   *Jun '19-May '20*
   o   Employed user research methodology for identifying social determinants of health (SDOH) in elderly community.
   o   Designed survey and interview questionnaires for collecting data corresponding to factors influencing SDOH.
   o   Applied statistical testing and regression modelling in R to predict falls from motion sensor data.

Trainee Decision Scientist
MuSigma Business Solutions Private Limited   *Oct '17-Mar '18*
   o   Developed ML, statistical testing, and visualization strategies for data from retail and aviation sector.

TEACHING/MENTORING EXPERIENCE

Community Mentor
Bioinformatics Research Network   *Mar '23-Present*
   o   Offer professional and scientific mentorship to bioinformatics enthusiasts from all parts of the world.
   o   Assist a high school student in understanding the basics of RNA-seq and scRNA-seq and guide them through the process of conducting differential expression (DE) analysis on samples from a rat cell line.

Facilitator
Genomics Scholars Program, Duke University   *Mar '21-Present*
   o   Develop resources introducing genomic and genetic concepts for participants in post-baccalaureate programs.
   o   Facilitate a hands-on session on Introduction to Genome Browsers with exercises on how to navigate IGV.

POSTER PRESENTATIONS

1. Shigaki D, Oh J W, Zhang J, Ko K, **Venukuttan R**, Barrera A, ENCODE FCC, Yu H, Beer M, Reddy T, Yardmici G, Siklenka K. Functional characterization of the mammalian genome. American Society of Human Genetics (ASHG) Annual Conference (2023). Walter E. Washington Convention Center, Washington D.C.
2. Clarke S, Beaman M, Duan Y, Iyengar A, **Venukuttan R**, Harris E, Ashley-Koch A, Gbadegesin R, Olabisi O, Majoros W, Reddy T, Ochoa A. Duke Genomics Scholars Program: Providing Accessible Genomic Training for a Diverse Workforce. NIH Centers for Excellence in Genomic Sciences (CEGS) Annual Meeting (2023). New York University, New York, NY.
3. Rai R, **Venukuttan R**, Barrera A, Allen A, Reddy T, Gersbach C, Crawford G. Nextflow Cellranger based pipelines for Perturb-seq experiments. The NHGRI Impact of Genomic Variation on Function (IGVF) Annual Meeting (2023). Washington University at St. Louis, St. Louis, MO.
4. Ochoa A, Duan Y, **Venukuttan R**, Clarke S, Reddy T. Development of Genomic Resource Modules for a Diverse Workforce. NIH Centers for Excellence in Genomics Science (CEGS) Annual Meeting (2022). Duke University, Durham, NC.

SKILLS

Programming languages: R, Python, Bash
Computing tools: HPC (SLURM), Google Cloud Platform (GCP), Azure
Workflow languages: CWL, Nextflow
Bioinformatics packages: STAR, Kallisto, Cellranger, MACS2, Bioconductor, Homer, MEME Suite, Bedtools suite, Deeptools, bamtools, samtools, bcftools
Genome Browsers: IGV, UCSC Genome Browser
Genomic Database: NCBI, GWAS, ENSEMBL, 1000Genomes, ENCODE
Methods & Concepts: Gene Set Enrichment Analysis (GSEA), Differential Expression Analysis (DESeq2, edgeR, limma-voom), Differential Peak Calling (MACS2, STARRPeaker, csaw, CRADLE), Dimensionality Reduction (t-SNE, PCA), Sequence Alignment (BWA, STAR, Bowtie), Clustering (hierarchical, DP_GP, K-means)

ACADEMIC PROJECTS

Gene Set Enrichment Analysis for Human Genome
- o Quantified the effect of drug Tamoxifen on breast cancer cell line from Affymetrix Human Genome chip using differential expression analysis (limma) and perform GO term on the DE genes.

RNA-seq analysis for Pasilla knockdown in Drosophila
- o Preprocessed a combination of RNAi and mRNA-seq data to identify the significantly affected genes due to Pasilla knockdown from an embryonic cell line in Drosophila using GSEA.

ChIP-seq analysis for identifying Peak calling in E. Coli
- o Indexed the reference genome file and mapped it to the reads followed by peak calling analysis (MACS2) to identify enriched regions from reads aligned uniquely to the reference genome for E. Coli.

Face recognition using Eigenfaces for human and chimpanzee faces
- o Developed a supervised algorithm in Python for prediction on unknown faces from a database of human and chimpanzee faces.
- o Compared the supervised learning technique to the traditional method of eigen-vectors.

Unsupervised learning on EEG signals
- o Performed various unsupervised machine learning techniques like clustering and Principal Component Analysis (PCA) on Intracranial EEG signals to classify epileptic and non-epileptic seizures.

CERTIFICATIONS

- o **Introduction to Genomic Technologies** – *The Johns Hopkins University/Coursera*
- o **Introduction to TensorFlow for Artificial Intelligence, Machine Learning and Deep Learning -** *DeepLearning.AI/Coursera*
- o **R programming** – *The John Hopkins University/Coursera*
- o **Google Analytics for Beginners** – *Google Analytics for Business*