Shrishtee Kandoi

(617) 838-1171 | shrishtee.kandoi@mssm.edu | https://www.linkedin.com/in/shrishtee-kandoi/

SUMMARY

I am a highly adaptable Bioinformatician with a strong foundation in computational human genetics and transcriptomic data. Additionally, I am proficient in high-performance computing, bioinformatics tools and software, genomic databases, clinical data management, data visualization, collaborative coding, automation, machine learning algorithms and genomic analysis on cloud platforms such as Google Cloud and Amazon Web Services.

SKILLS

Programming Languages: R, Python, Bash, Basic HTML and SQL

Tools and Software: bcftools, samtools, STAR, RShiny, Singularity conductors, Bioconductor, SoS Workflow system, GCP, AWS

Methods and Concepts: Gene set enrichment analysis, Gene Ontology enrichment analysis (DAVID), Differential Gene expression analysis (DESeq2, EdgeR), Dimensionality reduction methods (t-SNE, PCA, hierarchical clustering), Sequence Assembly & Alignment (BWA/STAR/BowTie), Statistical Analysis (Fisher's exact test, Permutation test, Student's t-test), Machine Learning Algorithms (Perceptron, K-means, K-Nearest Neighbor, Decision trees)

PROFESSIONAL EXPERIENCE

Icahn School of Medicine at Mount Sinai, New York, NY — Jan 2024 – Present

Bioinformatician I, ***Crary Lab*** *– Department of Genetics and Genomic Sciences*

- Actively contributing to the enhancement of data infrastructure, ensuring scalability, security, and seamless integration of diverse data sources for ongoing research and data science initiatives within the Lab.
- Spearheading the development of a robust pipeline focusing on running Genome-Wide Association Studies (GWAS) while ensuring Sample and SNP level QC and computing Polygenic Risk Score calculation, demonstrating expertise in data analytics and pipeline architecture.

Icahn School of Medicine at Mount Sinai, New York, NY — Jan 2023 – Jan 2024

Bioinformatician I, ***Sharp Lab*** *– Department of Genetics and Genomic Sciences*

- Leveraged bioinformatics tools such as STRetch and Expansion Hunter Denovo on cloud platforms like Terra and DNAnexus for genome-wide screens, identifying Tandem Repeat Expansions (TREs) in short-read genome sequencing that show causative associations for neurodegenerative diseases.
- Implemented quality control measures, including PCA analysis, to ensure data integrity and reliability followed by estimating size of these repeats by performing targeted search using a custom catalogue of potential pathogenic TR's.
- Conducted outlier detection and enrichment analyses, applying statistical methodologies including Logistic regression and Fisher's exact test to compare frequency of outlier expanded alleles in cases vs controls in the UK Biobank and All of Us dataset.

Boston University School of Medicine, Boston, MA — Jun 2022 – Dec 2022

Bioinformatics Summer Intern, **Zhang Lab** *– Biomedical Genetics Section*

- Experienced working with Linux environment, high performance computing clusters – Shared Computer Cluster (SCC)
- Performed next generation sequencing data analysis using data from the ADSP cohort - Religious Orders Study / Memory and Aging Project (ROSMAP) including quality control, sequence alignment, expression quantification, and multi-sample count normalization.
- Implemented routine quality control checks at various checkpoints and contributing scripts to the QTL analysis pipeline with additional metrics to be included in the next version of the pipeline.

EDUCATION

Boston University, Boston MA — Sep 2021 – Jan 2023

Master of Science in Bioinformatics

Earned merit-based scholarship

Relevant Courses: Foundations of Programming & Data Analytics, Machine Learning in Bioinformatics, Applications of Machine Learning in Systems Biology

Thapar Institute of Engineering and Technology, Patiala, India — Aug 2016 – Jul 2020

Bachelor of Technology, Biotechnology

Honors: Deans List

Relevant Courses: Immunology w/Lab, Biochemistry I & II w/Lab, Bioinformatics, Biostatistics, Computational Biology

ACADEMIC PROJECTS

Biological Database: Mutational Accumulation Data Aggregation Mar 2022 – May 2022

- Collaborated with teammates to develop an open-source web interface that relocates and pools raw data files, cross references feature and performs statistical aggregation to easily visualize queries against experimental data and automate workflows.
- Created Entity-Relationship diagrams to build a database focusing on identifying C to A mutational frequency in yeast strains containing *msh2* mutation
- Wrote python and SQL scripts to read and create tables of Whole genome sequencing data from multiple files and put them into a database.
- Utilized HTML-CSS and JavaScript to build the front-end and back-end of the database respectively.

Single Cell RNA-Seq Analysis of Pancreatic Cells Mar 2022 – April 2022

- Analyzed single-cell RNA sequencing in a set of post-mortem human donor pancreatic cells from four subjects and two mouse models to better understand the cellular diversity in the pancreas using Seurat-R Bioconductor package.
- Collaborated with research team members on experimental design and interpretation of results.
- Managed own time effectively across multiple projects while meeting deadlines set by supervisors/colleagues; prioritized work based on importance relative to other ongoing projects.
- Analyzed the UMI counts to identify clusters and marker genes for distinct cell type populations to ascribe biological meaning to the clustered cell types and identify novel marker genes associated with them.

Extrapolating transcriptional profile of murine cardiac regeneration using mRNA-Seq data Feb 2022 – Mar 2022

- Performed quality control on FASTQ files and recreated figures from O'Meara et. al., 2015
- Obtained mRNA-Seq datasets from GEO to generate estimates of transcript abundance/depletion in postnatal day 0 (P0) and adult (Ad) murine cardiomyocytes utilizing CuffLinks package.
- Compared FPKM counts of genes involved in three biologically distinct functional processes using hierarchical clustering algorithm and heatmaps to better interpret upregulation / downregulation between datasets.

Microarray based tumor classification Jan 2022 – Feb 2022

- Analyzed Microarray data and performed large scale genomic analysis including quality control, data preprocessing, running pipelines, downstream analysis, and result visualization.
- Developed a pipeline to process microarray data including normalization (RMA), batch correction, and scaling.
- Employed Principal Component Analysis to reduce dimensionality and examine outliers using ggplot2.
- Compared differentially expressed genes using GSEA package for KEGG, GO, and Hallmark collections.
- Maintaining best practices and using version controlling system (git) to maintain code.

LEADERSHIP AND TEAMWORK EXPERIENCE

Biotechnology Society of Nepal, Kathmandu, Nepal Jul 2020 – Jul 2021

- Organized bi-weekly webinars featuring guest speakers to bring students and communities together and educate them on pressing issues such as immunization and vaccination.
- Updated the E-bulletin section for COVID-19 vaccination and weekly case rates in local communities.
- Presented academic projects in the NRNA 2nd Global National Conference

Club Director, International Services, Rotaract Club, TIET Jun 2017 – Jul 2018

- Chaired the Rotaract Club and executed projects ranging from leadership training and workshops to blood drives and health camps.
- Facilitated the club treasurer to streamline a new budget system and quarterly forecasting process to receive funding from the Rotary club