

SHASHANK KATIYAR

github.com/shashkat linkedin.com/in/shashkat skatiya2@andrew.cmu.edu +1 878-834-9206

EDUCATION

Carnegie Mellon University

Master of Science in Computational Biology

GPA: 4.12/4.33

Pittsburgh, PA

December 2024

Indian Institute of Technology Kanpur

Bachelor of Technology in BioSciences and Bioengineering

GPA: 9.1/10 – Department Rank: 1

Kanpur, India

June 2023

SKILLS

Programming: Python, R, C/C++ , Golang, Shell Scripting (UNIX/Linux), SQL

Spatial Transcriptomics: Visium, VisiumHD, Xenium, cell2location, CellPhoneDB, Tangram, spatialdata

Single-Cell Transcriptomics: Scanpy, Seurat, Dropletutils, Scrublet, SoupX, Harmony, scvi-tools

Bioinformatics: Cellranger, Spaceranger, Nextflow, FastQC, UCSC Genome Browser, Salmon, bedtools

Toolkit: Pandas, Numpy, Matplotlib, Seaborn, Pytorch, Git/Github, Docker, AWS Sagemaker, HPC

EXPERIENCE

Novasenta Inc.: Cancer Immunology

Computational Biology Intern

Pittsburgh, PA

May 2024 – August 2024

- Identified celltype abundances in Lung Cancer Visium samples using cell2location spot deconvolution, improving company's understanding of localization of target ligand-receptor (L-R) pairs in tumor tissue
- Inferred celltype-celltype interactions using Linear NCEM (Node Centric Expression Models) in spot deconvoluted samples gaining further evidence of the L-R interaction between expected celltypes
- Validated deconvolution by performing scRNAseq data integration with spatial data using Tangram
- Quantified L-R colocalization in VisiumHD data with binomial test; got significant colocalization, $p < 0.01$
- Implemented Nuclei Segmentation on VisiumHD data to get spatial-single cell data for L-R analysis and found multiple celltype-pair interactions with pvalue < 0.05 according to custom-made permutation test

TenSixty BioSciences: Cancer Drug Discovery

Machine Learning Intern

Cambridge, MA

Sep 2022 – July 2023

- Designed pipeline for Marker Gene Identification of rare cell types found in cancer, employing NGS scRNA sequencing data and Scanpy Library, and discovered their prevalence in Cancers from TCGA
- Collaborated with Amazon Web Services officials to execute Kallisto sequencing tool on Amazon Web Services parallel, with Docker, for high throughput processing of raw FASTQ files at isoform level
- Identified a differentially expressed isoform for small cell lung cancer for specific cancer drug targeting
- Designed a protein binder for identified isoform with RFdiffusion with Alignment Error of 10

PROJECTS

Single Cell Analysis of Engineered AAVs for Gene Therapy

Mentor: Prof. Leah Byrne, University of Pittsburgh

Pittsburgh, PA

Jan 2024– May 2024, Aug 2024 – Present

- Implemented Slurm scripts to run Cellranger on HPC cluster on scRNAseq FASTQs from primate retina
- Performed quality control followed by celltype annotation on processed scRNAseq files using Scanpy
- Identified presence of different adeno-associated virus (AAV) barcodes in cells using Salmon tool
- Mapped AAV infection patterns to celltypes, revealing preferences of AAV variants for cell populations
- Developing Nextflow workflow for the processing pipeline of quantification of AAV-barcodes in cells

Gene Expression Prediction Using Deep Neural Network

Self Project

Pittsburgh, PA

December 2023

- Developed Nextflow pipeline to process tissues' RNAseq & integrate with ATACseq data from same tissue
- Implemented Deep Neural Network using ATACseq peaks as input for accurate estimation of gene expression, obtaining a Mean Squared Error of 0.032 between predicted and target expression values

PUBLICATIONS

Vector Database of COVID-related papers from CORD-19

Deep Learning, Biomedical Data and Knowledge Graphs Hackathon by DNaNexus

Pittsburgh, PA

Oct 2023

- Developed a Python pipeline to build a custom vector database from the CORD-19 dataset and pre-trained LLM and queried it successfully for specific contexts to yield relevant scientific articles