Shizhao Yang

+1 206 688 9210| syang71@uw.edu |GitHub| LinkedIn

SUMMARY OF QUALIFICATIONS

Experienced biostatistician with a strong foundation in data science, focusing on genomics and statistical modeling, with a proven track record in RNA-seq analysis, metagenomics, and bioinformatics. Skilled in leveraging biostatistical methods for genomic data interpretation, algorithm optimization, and predictive modeling in disease research. Strong problem solving and collaboration abilities; excellent written, verbal and visual communication skills.

EDUCATION

University of Washington, Seattle, WA *-expected Mar 2025*

M.S. in Biostatistics

New York University, New York, NY *May 2023*

B.S. in Data Science, Genomics concentration (Minor: Mathematics)

WORK EXPERIENCE

Python-based RNA-seq Analysis Algorithm using Negative Binomial GLM New York, USA, *June-Dec 2022*

Research Assistant *Supervisor: Manpreet Katari, New York University*

- Built a python-based Stats Model based on Generalized Linear Model
- Applied backtracking line search and IRLS in coefficient estimation and Wald Test to the estimated log fold changes.
- Optimized runtime using Python multiprocessing and validated against Deseq2 data.

Investigation of Horizontal Transfer in Metagenomics Shanghai, China*, June 2021-Jan 2022*

Research Assistant *Supervisor: Gang Fang, NYU Shanghai, NYU*

- Conducted RNA-seq analysis of human gut microbiome data using Shell, encompassing genome assembly, mapping, gene-calling, and annotation.
- Utilized the Louvain Method for creating pseudo ortholog communities and calculated TPM and a self-defined PI index to study gene persistence.
- Analyzed differential gene expression related to Horizontal Gene Transfer using statistical methods like ANOVA, correlating TPM and PI using Python and R.

PROJECT WORK

Refined SIR Model with Vaccination and its Application in 2022 NYC Influenza A Activity Prediction

- Analyzed a modified SIR model incorporating vaccination using ODE methods like fixed point stability, phase plane, and herd immunity.
- Simulated six years of NYC influenza data with the SIRV model, estimating transmission and removal rates, and R0 using the Quasi-Newton method in Python.
- Compared past season coefficients to predict this year's NYC infection peak under various vaccination rates.

LANGUAGES & PROFESSIONAL SKILLS

- Chinese (native), English (fluent)
- Programming Languages: Python, R, Shell, SQL, MongoDB, HTML, MATLAB, Javascript
- Analytics Skills: Statistical modeling, Numerical analysis, Machine Learning
- Bioinformatics Applications: SPAdes, Bowtie2, Hisat2, Samtools, diamond, Blast, Limma, Deseq2