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 (Modeling and Methods pathway)
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*
- Developed a Python-based statistical model with GLM framework for RNA-sequence analysis, which incorporated backtracking line search and IRLS for coefficient estimation and the Wald Test for assessing log fold changes.
- Reduced the algorithm's performance time by 80% through the utilization of Python multiprocessing, and verified its reliability with the DESeq2 package.

Investigation of Horizontal Transfer in Metagenomics   Shanghai, China*, June 2021-Jan 2022*
Research Assistant   *Supervisor: Gang Fang, NYU Shanghai, NYU*
- Executed comprehensive RNA-seq analysis of human gut microbiome through shell for genome assembly, mapping, and annotation; employed the Louvain Method to generate pseudo ortholog communities for gene persistence studies.
- Analyzed differential gene expression related to Horizontal Gene Transfer using statistical methods like ANOVA, correlating TPM and gene persistence using Python and R.

PROJECT WORK

Single-cell Lineage Embedding Contrastive Learning   Seattle, WA, *Nov 2023-present*
Researcher   *Supervisor: Kevin Lin, University of Washington*
- Designed and implemented a contrastive learning algorithm to learn the high-dimensional embeddings of single-cell data, facilitating the identification of lineage-specific gene expression patterns.
- Develop an evaluation metric for comparing the different embeddings of single-cell data generated by the model.

Refined SIR Model with Vaccination and its Application in 2022 NYC Influenza A Activity Prediction
- Employed a modified SIR model to analyze vaccination effects on influenza spread in NYC, utilizing ODE methods for fixed point stability and herd immunity assessment.
- Leveraged the SIRV model to simulate NYC influenza trends over six years, calibrating transmission dynamics and forecasting peak infection rates for the year 2022 based on historical data and vaccination scenarios, using the Quasi-Newton method in Python.

LANGUAGES & PROFESSIONAL SKILLS

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