RAEHASH SHAH

+1 (201) 449-4918 | raehashshah@gmail.com | PORTFOLIO: https://shahraehash.github.io | GITHUB: https://github.com/Shahraehash

EDUCATION

Integrated Master of Science in Computational Biology   2024 – 2025

School of Computer Science, Carnegie Mellon University // Pittsburgh, PA

- *Thesis: "BatchBlend: A Batch-Aware Interpretable Probabilistic Graphical Model for Integrating Spatially Resolved Transcriptomes"*

Bachelor of Science in Computational Biology *University Honors*   2020 – 2024

School of Computer Science, Carnegie Mellon University // Pittsburgh, PA

Bioinformatics Tools (in Course projects): BLAST, Bioconductor, Clustal, SAMtools, Seurat, GEO, GATK, VMD, PyMOL, Galaxy
Machine Learning Tools (in Course projects): Conda, Git, Pandas, Pyspark, Pytorch, Tensorflow, Kaggle, Google Colab, AWS

EXPERIENCE

Computational Biologist at GEn1E Lifesciences // Palo Alto, CA   10/2023 – 08/2024

- Classified diseases in Python using publicly accessible data based on clinical profiles, molecular pathways, gene expression profiles and epidemiological data to identify diseases similar to the company's current drug target profile.
- Created a PubMed search engine that accesses all articles relevant to a search query and uses Natural Language Processing & Knowledge Graph techniques to characterize and summarize the results of those articles which accelerates the time it took for a scientist to find published work and data from a search query by over 25%.

Bioinformatician at Regeneron // Rensselaer, NY   05/2023 – 07/2023

- Trained a Structured State Space Diffusion model on Oxford Nanopore Sequencing Technology signal data in a Good Manufacturing Processes environment to generate viral and host signal data. The generated signal data was then passed through a machine learning model, basecaller, to convert it to DNA sequencing data. The generated and true DNA reads followed a similar distribution in sequence length and were indistinguishable when evaluated on a binary classifier that was trained on both reads.

Software Systems Engineer at Werfen // Bedford, MA   05/2022 – 08/2022

- Designed and developed a Time-division multiplexing Protocol Sniffer for a RS485 Bus using a Docker environment to visualize the packets of information that were being passed between the layers of a throughput hemostasis testing instrument firmware.
- Led daily standups with a cross-functional team as a Scrum master in an Agile work environment (Agile Atlassian Certified).

LEADERSHIP

Teaching Assistant for Courses at Carnegie Mellon University   01/2023 - 05/2024

- Courses: Machine Learning in Health Care, Computational Medicine, Great Ideas in Computational Biology
- Organized course material, created problem sets, taught recitations & answered students' questions to promote learning in courses.

President of Computational Biology Undergraduate Society at Carnegie Mellon University   09/2022 – 05/2024

- Spearheaded the creation of the Integrated Master's program in Computational Biology by assisting in designing the curriculum and obtaining administrative approval.
- Co-Organizer of International Society for Computational Biology GLBIO 2024 Conference Workshop.
- Chaired 6 workshops each year connecting students with experts in research and industry, fostering opportunities for collaboration.

PUBLICATIONS

"Pan-Cancer PDOs Preserve Tumor Heterogeneity and Uncover Therapeutic Vulnerabilities" (2025) Kuo H., Bhinder B., … **Shah R.,** … Elemento O., Sboner A., Martin M. L., https://doi.org/10.1101/2025.04.10.647635 (Journal publication pending)

- Analyzed tumor heterogeneity across 300 primary samples and organoids at different points of cell culture through Hierarchical Clustering, PCA and Differential Gene Expression Analysis in R.

"*SCN9A*: Proposal of Voltage-Gated Ion Channels as a Novel Diagnostic Marker for Alzheimer's Disease" (2023) Shah D., **Shah R**., Waldron A., Leonardi D., https://doi.org/10.1101/2023.05.18.23289925

- Performed Gene Set Enrichment Analysis and Logistic Regression Analysis on mRNA and bulk RNA-seq data.

SKILLS

Extracurriculars: Scuba Diving Certified, Soccer (Co-Captain), Swim Instructor, DDN Bollywood Dance Team, Hackathons
Programming Languages: Python, R, C, Java, HTML, Go, MATLAB | **Spoken Languages:** English, Gujarati, Spanish, Hindi