

Xinyi SHANG

413-695-4282 | xs2529@cumc.columbia.edu | <https://xinyishang.github.io> | [linkedin.com/in/xinyi-anna-shang](https://www.linkedin.com/in/xinyi-anna-shang)

EDUCATION BACKGROUND

Columbia University, Mailman School of Public Health
M.S. in **Biostatistics**, Theory and Method Track
GPA:4.00/4.00

New York, NY
09/2023-Exp 05/2025

Brandeis University
B.S. in **Biology** with Honors; Minor in **Computer Science**
GPA: 3.83/4.00; *summa cum laude*; **Dean's List** 2020 - 2023

Waltham, MA
09/2019-06/2023

ACADEMIC EXPERIENCE

Identifying Cell-type-specific Spatially Variable Genes

06/2024-Present

Supervisor: Prof. Wenpin Hou, Prof. Zhicheng Ji

- Integrated genomic and spatial data to identify spatially variable genes within specific cell types in spatial transcriptomes, aiding in the understanding of regional biological processes and functions.
- Assisted in annotating cell types using 10x Visium HD datasets from various species and tissues. Explored genetic and cell type databases such as Panglao DB and CellMarker 2.0 to gain insights into cell type-specific gene expression.
- Evaluated algorithms and analyzed ontology terms to enhance the understanding of gene expression patterns and their spatial distribution, contributing to the analysis of gene function and spatial heterogeneity in organs and tissues.

GeneTuring tests GPT models in genomics

08/2024-Present

Supervisor: Prof. Wenpin Hou, Prof. Zhicheng Ji

- Conducted thorough assessments of evaluating advanced GPT models (GPT-3.5, GPT-4, GPT-4o, Gemini Advance) across 1,200 genomics-related questions, benchmarking their performance on tasks such as gene extraction, SNP location, and gene-disease association, protein coding, gene ontology, gene alias association, etc.
- Analyzed model accuracy by focusing on the models' ability to recognize task limitations and handle unknown queries.

Thalamic Nuclei Derived Radiomics & Volumetric Trajectories in MS

05/2024-06/2024

Supervisor: Dr. Korhan Buyukturkoglu

- Analyzing radiomic and volumetric data to uncover the relationship between thalamic nuclei changes in Multiple Sclerosis (MS) patients and cognitive impacts measured by the Symbol Digit Modalities Test (SDMT).
- Processed and analyzed data from 126 MS patients, focusing on 8 bilateral thalamic nuclei, the whole thalamus, and over 1,500 radiomic features. Identified 2 distinct MS subtypes based on volumetric and 2 subtypes using radiomic data.

Senior Thesis: Discrimination in An Attractor State Network with Short-term Synaptic Plasticity

06/2022-05/2023

Supervisor: Prof. Paul Miller

- Study the short-term synaptic plasticity's effect in the discrimination of different stimuli sequences in a randomly connected attractor network. Perform analysis with a computational neuroscience model.
- Performed clustering and confusion matrix to investigate the networks' ability to discriminate with different parameter sets.
- Results showed that the presence of facilitation increased the network's stability and depression made the network more sensitive to different signals. The network with both facilitation and depression demonstrated the most complex behavior.

AWARDS

Grand Prize Winner, Hacking Health Hackathon 2024, Columbia University

02/2024

Silver Medal, Kaggle, Competition: HMS - Harmful Brain Activity Classification

02/2024-04/2024

Semi-finalist, Women's Health Tech Challenge at HITLAB, AI4Purpose Inc.

08/2024-09/2024

Manuscript & POSTER ABSTRACT

Zhuang, H., **Shang, X.**, Hou, W., & Ji, Z. (in progress). Identifying Cell-Type-Specific Spatially Variable Genes with ctSVG2.

Buyukturkoglu, K., Davis, L., Wen, S., **Shang, X.**, Zhang, W., Comandate-Lou, N., Blackwelder, J., Shende, V. K., Ozcelik, S., Boulanger, A., Riley, C., Stern, Y., & De Jager, P. (2024). Thalamic Nuclei Derived Radiomics and Volumetric Trajectories in Multiple Sclerosis and Their Associations With Symbol Digit Modalities Test. Presented at the 40th Congress of ECTRIMS, Copenhagen, Denmark.

TEACHING EXPERIENCE

Teaching Assistant, Applied Regression II, Columbia University

09/2024-Present

Teaching Assistant, Statistical Computing with SAS, Columbia University

09/2024-Present

Facilitator, Summer Health Professions Education Program (SHPEP), Columbia University

06/2024-07/2024

WORK EXPERIENCE

R&D Intern, Tiangen Biotech (Beijing) Co., Ltd

03/2021-05/2021

SKILLS

- **Programming Language:** Python, R, Matlab, Java, SAS (Base & Advanced Certified), SQL
- **Machine Learning & Deep Learning:** Regression Analysis, Classification, Clustering, Neural Networks, Computer Vision
- **Biostatistics & Statistical Modeling:** Survival Analysis, Longitudinal Analysis, Clinical Trial Design, Linear Mixed Model