2024 Summer Scholar Profile: Alice Zhang

My name is Alice Zhang, and I am a rising senior at Wellesley College majoring in Data Science with a concentration in Bioinformatics. At school, I work in the Kellis Lab under Professor Manolis Kellis, where I research inclusive polygenic risk score models (iPGS) and the systematic application to brain MRI-derived traits with my mentor Dr. Yosuke Tanigawa. My project involves using iPGS to analyze brain MRI-derived traits, such as brain volumetric measurements. The project is scaling up across many brain MRI-derived traits, integrating with brain eQTL datasets, and modeling multiple response variables/brain MRI-derived traits simultaneously. At the Buck Institute, I interned in the Schilling Laboratory under Professor Birgit Schilling with my mentors Dr. Joanna Bons and Dr. Mark Watson. The Schilling Lab studies the molecular mechanisms involved in aging and age-related diseases through uncovering protein signatures and pathways using mass spectrometric technologies.

My first project was developing a protocol to help researchers generate data-dependent acquisition (DDA) protein posttranslational modification (PTM) spectral libraries for proteome-wide lysine succinylation analysis of data-independent acquisition (DIA) data. DIA simultaneously collects fragment ion spectra for all detectable precursor ions isolated within a defined wide m/z range, generating highly complex spectra with fragment ions from multiple co-eluting peptides and allowing for comprehensive, reproducible, and unbiased quantitative proteome analysis. Deconvoluting DIA MS/MS spectra is, however, computationally intensive, requiring complex, sophisticated and dedicated analysis tools. One approach is to use the information contained in a spectral library generated from DDA data to perform DIA data extraction, peak group scoring, and PTM site localization probability determination for peptide and PTM identification and quantification. DDA relies on iteratively selecting precursor ions for fragmentation based on predefined criteria, typically ions with the strongest signals from the survey scans. Collected DDA MS/MS spectra are then submitted to database search for identification. A spectral library generated from high-quality DDA data contains experimentally observed MS/MS spectra matched to peptide sequences with high confidence, leading to high identification confidence, sensitive and accurate quantification, and enhanced probabilities of correctly identifying and accurately localizing the modifications, which is necessary to understand their biology. Using DDA data collected on mouse brain succinyl enrichments, we generated spectral libraries in Spectronaut (Biognosys), Spectromine (Biognosys), Proteome Discoverer (Thermo Fisher Scientific) and MSFragger through FragPipe (Nesvizhskii Lab), which are all search engines, applying finely optimized settings. We reported our results and conclusions in the protocol paper, “Generating High-Quality Data-Dependent Acquisition Succinyl Spectral Libraries for Proteome-wide Succinylome Analysis by Data-Independent Acquisition” to be submitted in *Methods in Molecular Biology*.

Cellular senescence is hypothesized to be involved in aging and many age-related diseases. The Schilling laboratory is very collaborative, and my second project was in collaboration with the Campisi laboratory at the Buck Institute, working on the R Shiny app FAST.R. FAST.R is part of the Fully-Automated Senescence Test (FAST) workflow developed by Dr. Francesco Neri that allows for a relatively simple way of automated detection of cellular senescence (Neri *et al.*, *GeroScience*, 2024). Because there is no universal marker for senescence, it is typically determined manually through looking at several senescence associated markers. FAST was developed and optimized to automatically and rapidly detect senescence in cell cultures with a predefined set of three markers; however, we wanted to make FAST compatible with any cell types. I spent the second half of my summer recoding FAST.R, adding flexibility to the number and types of markers, and published the new code to GitHub. We have tested this code on cells from the brain and cartilage, demonstrating the flexibility, utility, and ease-of-use of this new iteration of FAST.R. This code will be used in future FAST projects that are being developed in the lab now.