Emily Liang

0988535798 | emilywliang@gmail.com | Taipei, Taiwan

EDUCATION

University of California, San Diego San Diego, CA
B.S. in Bioinformatics, minor in Computer Science - GPA: 3.66/4.00 *Sep. 2020 – June 2024*

Relevant Coursework: Probability and Statistics, Data Science in Practice, Molecular Sequence Analysis, Data Structures and Object-oriented Design, Design and Analysis of Algorithms, Applied Genomic Technologies

RESEARCH & PROJECTS

Research Lab Volunteer Nov. 2024 – March 2025
Tukey Lab, UCSD *La Jolla, CA*
- Conducted PCR assays to amplify DNA for analysis, ensuring accuracy in sample preparation and handling.
- Collected and processed tissue samples from mice, following standardized protocols for tissue preservation and documentation.
- Assisted with routine laboratory tasks, including reagent preparation, sample labeling, and equipment maintenance, to support daily lab operations.
- Maintained cleanliness and organization of lab areas, adhering to strict safety and quality control standards.

Tumor Tissue Type Analysis for ecDNA and Gene Identification Jan. 2024 - March 2024
Python, BLAST *Team size: 3*
- Analyzed 500+ ecDNA regions to identify 30+ key tumor-driving genes.
- Optimized BLAST workflows, reducing search time by 40%.
- Implemented efficient filtering processes to pinpoint ecDNA regions lacking high-frequency genes, enabling a focused analysis of potential driver genes.

Protest-Related Lexicons and Magnitude of Protests Data Analysis March 2024 - June 2024
Python, TF-IDF, Vader *Team size: 5*
- Analyzed 6 protest-related lexicons; identified 2 key terms ("protest" and "riot") with significant correlations to protest sizes ($r = 0.62$ and 0.32).
- Conducted OLS regression; achieved statistically significant results ($p = 0.0$) for key terms predicting protest density.
- Quantified word frequency trends across protests using TF-IDF, linking increases in specific lexicons to social movement magnitudes.
- The project provided actionable insights by identifying significant lexicons ("protest" and "riot") correlated with protest magnitudes, advancing the understanding of social movement dynamics and demonstrating the utility of TF-IDF and OLS regression for such analyses.

KmerEstimator: Estimating Optimal K-mer Sizes for Genomic Analyses March 2024 - June 2024
Python *Team size: 3*
- Developed a Python program inspired by KmerGenie to estimate optimal k-mer sizes for genomic datasets.
- Implemented workflows to preprocess sequence data and calculate k-mer histograms using custom algorithms.
- Designed statistical methods to evaluate k-mer distributions and identify optimal sizes for assembly quality.
- Gained hands-on experience with algorithm design, sequence data analysis, and benchmarking against existing tools.

TECHNICAL SKILLS

Programming Languages: Python, Matlab
Scripting and Tools: Bash, Linux, Github
Data Analysis: Statistical Methods (Descriptive, Inferential, Hypothesis Testing), Regression, Multivariate Analysis
Bioinformatics Tools: Sequence Alignment (BWA-MEM, STAR), Variant Analysis (VARSCAN, BCFTools), Quality Control (FASTQC, Sickle), Functional Analysis (BLAST, HMMER), Transcript Quantification (RSEM), Data Indexing (Jellyfish)
Biological Techniques: qRT-PCR, PCR, Gel Electrophoresis, RNA Extraction, Western Blotting