

Bioinformatics Developer with experience developing diagnostic and experience with building production diagnostic algorithms and data platforms. Skilled in Python backend development, Python visualization tools, and translating complex biological data into scalable software solutions.

Experience

Stanford University, Snyder Lab – Research Assistant

Oct 2025 – Present

Research Assistant contributing to host response research involving beneficial exposomes (cytokine responses to environmental and lifestyle exposures), single-cell spatial, and single-cell sequencing data.

- Developing Python-based host-response analysis pipelines for single-cell and cytokine data that measure longitudinal changes within cell lines and biological marker pathways using PCA, principal curves, and pseudotime modeling to identify both time and dose-sensitive response patterns across treatment concentrations.
- Unsupervised clustering and dimensionality reduction algorithms (PCA, UMAP) was implemented to track expression modulation and divergence within cell lines and expression data.
- Engineering data pre-processing workflows using pandas for multi-dimensional datasets including analysis of 45+ biomarkers for cytokines, enabling robust statistical analysis of dose-dependent immune responses.

Kelliop Inc. – Bioinformatics Developer

Jul 2021 – Present

Bioinformatics Developer contributing to in vitro diagnostic platforms for neonatal care, pharmacogenomics, and oncology using lab-on-chip technologies and advanced sequencing pipelines.

- Developed DNA analysis algorithms using bioinformatic protocols for 2 neonatal ICU and 1 endometrial cancer diagnostic applications; improved multiplexing accuracy by 15%. All three applications entering clinical trial with promising results.
- Represented company at four ADLM conferences; led product demos and technical discussions with 40+ clinicians and stakeholders.
- Delivered technical demonstrations of DNA sequencing workflows for Lab-On-Chip IVD devices, enhancing clinical adoption of diagnostic tools.

Education

University of California, Santa Cruz

B.S. in Biomolecular Engineering and Bioinformatics – Bioinformatics Concentration, Minor in Computer Science

Projects

Human Bimolecular Atlas Project (HuBmAP) - Intestine Data Analysis

- Analysis of single-cell spatial and RNA sequencing data for gastrointestinal tract and liver data
- Allows for more robust host response tracking with longitudinal data able to be used for pathology, drug development, identification of irregularities, and drug re-purposing analysis at a single-cell level.

Interactive Cytokine Trajectory Analysis Dashboard – Full-Stack Bioinformatics Platform

- Built production-quality platform combining Python backend (Flask, pandas, scikit-learn, statsmodels) with JavaScript for cytokine research.
- Performs bioinformatic analysis for dimensionality reduction (PCA, UMAP), statistical testing, and trajectory modeling endpoints.
- Implemented ComBat batch correction pipeline reducing technical variation by 70-90% across multi-study datasets, enabling valid cross-study comparisons.

Allergen Detection ML Pipeline

- Developed allergen detection pipeline using Random Forest Regression to model serum IgE levels from NHANES and ImmPort datasets.
- Implemented comprehensive data engineering workflows to integrate and clean multi-source clinical datasets; analyzed feature impact using SHAP.
- Built classifiers to identify allergenic genes/proteins using genomic data and Bayesian models.

Encrypted Data Structures for Secure Search

- Collaborated with team through Git/GitHub to implement oblivious RAM protocols (Path ORAM and rORAM) in C++ for secure storage systems.
- Benchmarked I/O performance across database sizes (2^{14} – 2^{16} blocks); reduced range query overhead by 300×.

Certifications

AWS Cloud Certified Practitioner

Amazon Web Services

Oct 2025

Skills

Programming & Development: Python, JavaScript (ES6+, D3.js), C/C++, SQL, Java, Bash, Git/GitHub, Linux/Unix

Machine Learning & Statistics: PyTorch, Scikit-learn, Pandas, NumPy, Bayesian Statistics, Dimensionality Reduction (PCA, UMAP, t-SNE), Statistical Testing (Mann-Whitney, Kruskal-Wallis, FDR correction)

Bioinformatics: Genomics, Sequence Analysis, UCSC Genome Browser, BLAST, GTEx, Batch Effect Correction (ComBat), Trajectory Analysis, Pseudotime Modeling

Data Engineering: Data pipeline development, Multi-study integration, Data normalization, Performance optimization

Professional: Technical Documentation, Cross-functional Collaboration, English (Fluent), Mandarin Chinese (Fluent)