

# Jonathan M. Anzules

Merced, CA, 95340

📞 646-712-1764    ✉ [jonanzule@gmail.com](mailto:jonanzule@gmail.com)    🔗 [linkedin.com/in/jonathan-anzules](https://www.linkedin.com/in/jonathan-anzules)    🐙 [github.com/J-Anzules/](https://github.com/J-Anzules/)

## Summary

I am a computational biologist with expertise in multi-omics analysis, specializing in developing statistical and machine-learning models to solve complex biological problems. My experience encompasses working at notable institutions such as Lawrence Livermore National Laboratory and the Joint Genome Institute, where I focused on leveraging cloud and high-performance computing resources to enhance bioinformatics analyses for large-scale studies.

## Education

- |  |                      |
|--|----------------------|
| <b>University of California, Merced</b><br><i>Ph.D. in Quantitative and Systems Biology</i>  | 2023<br>Merced, CA   |
| <ul style="list-style-type: none"><li>• <b>Project:</b> Collected and prepared autoimmune disease datasets, and developed a predictive mathematical model to replicate the observed data, aiming to enhance understanding and treatment strategies</li></ul> |                      |
| <b>Brooklyn College</b><br><i>Bachelors of Science - Biology</i>   | 2016<br>New York, NY |

## Experience

- |   |                                |
|---|--------------------------------|
| <b>PhageNet</b><br><i>Genomic Data Scientist (Startup / Freelance)</i>  | September 2021 – February 2024 |
| <ul style="list-style-type: none"><li>• Conducted pseudotime analysis of publicly available scRNA-seq data to reveal critical expression shifts between healthy and diseased beta cells in type 2 diabetes (T2D), identifying two dedifferentiation pathways.</li><li>• Conducted in-depth analysis using generative pre-trained transformer models, scGPT and scBERT, for automated cell type annotation and gene network perturbation studies in scRNA-seq data of T2D islet cells.</li><li>• Analyzed bulk RNA-seq and proteomic data of T2D islet cells, employing differential expression and gene set enrichment analyses to highlight distinct transcriptomic changes and pathways.</li><li>• Implemented a machine learning-based pipeline in Python, Perl, bash, and R to detect phylogenetically unique genomic prophage in Staphylococcus aureus genomes; identifying 196 novel genomes.</li></ul> |                                |
| <b>University of California, Merced</b><br><i>Biostatistician Scholar</i>   | August 2017 – May 2023         |
| <ul style="list-style-type: none"><li>• Utilized a flow cytometer to gather, statistically analyze, and quantify immune cell populations from 60 healthy and autoimmune mice (BALB/c IL-2 KO).</li><li>• Developed an ODE-based optimization pipeline to simulate the progression of autoimmune diseases, accurately modeling 8 distinct immune cell populations as observed in our dataset.</li><li>• Executed over 500,000 simulations using advanced statistical techniques to pinpoint crucial parameter spaces essential for autoimmune disease prevention.</li></ul>  |                                |
| <b>Lawrence Livermore National Lab</b><br><i>Proteomic and Genomic Data Scientist</i>   | May 2022 – August 2022         |
| <ul style="list-style-type: none"><li>• Curated and hosted a MySQL database on AWS, comprising 3,000 SARS-CoV-2 genomes and proteomes, for advanced analysis using the transformer-based machine learning model ESM by Facebook Research.</li><li>• Engineered a pipeline leveraging ESM to identify viral variants and predict protein function/fitness; achieving an accuracy rating of 67%.</li><li>• Led a team of four data scientists in the development and implementation of a random forest and graph neural network model of protein interactions, successfully identifying novel antiviral candidates against SARS-CoV-2; achieving an accuracy rating of 74%.</li></ul>   |                                |
| <b>Teaching Assistant for Natural Sciences</b><br><i>Math015</i>  | August 2017 – May 2023         |
| <ul style="list-style-type: none"><li>• Directed a class of 30 students, facilitating assignments and discussions that introduced them to statistics and R programming.</li><li>• Created labs that are approachable and encourage excellent coding practices.</li></ul>  |                                |

## SIAM Applied Math Challenge - UC Merced

March 2021 – May 2021

### Team Lead

- Mentored a group of 4 students in the development of a SIR model in Python.
- Conducted a sensitivity analysis on parameters of interest.
- Received second place in model solution and presentation.

## Lawrence Livermore National Lab

May 2019 – July 2019

### Team Lead for the LLNL Data Science Challenge

- Led a team of 5 students in the development of a machine learning algorithm in Python to improve the antibody binding affinities.
- Improved model's predictive performance by identifying the most critical features in our dataset.

## Department of Energy - Joint Genome Institute

July 2017 – August 2017

### Genomic Bioinformatician Intern

- Explored Bayesian change point analysis for identifying chimeric sequences generated during the library prep step for sequencing by Oxford Nanopore MinION.
- Reviewed code associated with the handling of data produced by the Oxford Nanopore MinION.
- Learned industry standards necessary for collaborating with code scripts, data, and cluster usage.

## Publications

---

- **Anzules, Jonathan Michael.** "Mathematical Modeling of IL-2 Dysregulation in the Pathogenesis of Autoimmune Disease." Dissertation, University of California, Merced, 2023.
- **Jonathan M. Anzules**, Lihong Zhao, Kristen M. Valentine, Genevieve N. Mullins, Katrina K. Hoyer, Suzanne S. Sindi. "Mathematical Modeling of homeostatic expansion in healthy and autoimmune system" *Journal of Immunology*, expected submission May 2024.
- Madelyn C. Houser, **Jonathan Anzules**, Tyrome Sweet, Loukia Lili-Williams, Steven Bosinger, T. Grauman, Khaled Machaca, Dean Jones, Young-Mi Go, Vicki S. Hertzberg, Susan Safley, Collin J. Weber. "Transcriptomic, Metabolomic, and Proteomic Analyses of Type 2 Diabetic and Non-Diabetic Pancreatic Islets." Preprint.

## Workshops Mentored

---

- February 2022 - scRNA-seq Workshop at UC Merced
- March 2019 - UC Merced Research Viz Workshop
- June 2018 - Python workshop at UC Merced

## Data Analytics Skills

---

**Programming Languages:** Python, R, SQL, Matlab, Bash, Perl, Julia

**Software & Tools:** Scikit-learn, PyTorch, HPC clusters, DESeq2, Salmon, Seurat, Scanpy, Muon, HISAT, PhiSpy, BLAST, FastQC, UMI-tools, Samtools, Bioconductor, Anaconda, Matplotlib, ggplot, Seaborn, Jupyter, Pandas, NumPy, SciPy, Limma-Voom, Bowtie, Tophat, PROKKA, MUSCLE

**Organization Tools:** Trello, Slack, Amazon Web Services

**Language Fluency:** English, Spanish