Joshua P. Schaaf

610-550-6432 • Davis, CA • joshuaschaaf@gmail.com • https://github.com/viabard

SKILLS

Python | R | Git | SQL | C | JavaScript | Linux | Excel | PowerPoint | Pandas | NumPy | scikit-learn | TensorFlow Keras | PyTorch | Matplotlib | PCR | Methylation | FASTA/Q | ChIP-Seq | Proteomics| Github Actions | AWS

EXPERIENCE

Bioinformatics Programmer Analyst II, HJF WHIRC GYN-COE, Remote *January 2023-Present*

- Develop, publish (co-first author), and deploy **ProteoMixture**, a method of determining relative abundances of different tissue types (stromal, tumor, immune) within a bulk tumor tissue sample. Tissue-specific gene sets were determined using samples with known ratios of tumor, stromal, and immune cells by performing differential abundance analyses with *limma*, and then I applied a machine learning (ML) method (recursive feature elimination using support vector machines with linear kernels) to refine gene sets further. https://lmdomics.org/ProteoMixture/
- Analyze, design, develop, modify, test, and implement software tools/scripts (R, Python, bash) supporting bioinformatic analysis of proteomic, transcriptomic, genomic, and clinical data
- Package and present bioinformatic analysis results of proteomic (PCA, differential abundance, supervised and unsupervised clustering algorithms, GSEA, GO, survival) for external collaborators
- Develop ML models with both proteomic and clinical data for research and clinical application, utilizing multiple feature reduction approaches to determine highly predictive biomarkers

Data Analyst, FOXO Technologies, Remote *July 2021-November 2022*

- Analyzed 'wide' epigenetic data (~1 million features) using Python (pandas, numpy) and R to provide accurate predictors for clinical targets using automated machine learning (ML) platform DataRobot. Analysis using working knowledge of *scikit-learn* and other ML/data mining algorithms
- Created easy to understand figures/interactive dashboards with data visualization tools in Python (plotly, matplotlib, seaborn), R (ggplot), and lucidchart
- Performed Quality Control and Validation analyses on big data, utilizing Python in AWS Sagemaker Processes, Sagemaker Notebooks, Cloudwatch, storing data on AWS S3
- Developed RAPA (Robust Automated Parsimony Analysis), a Python package for performing recursive feature reduction of auto-ML models on DataRobot https://github.com/FoxoTech/rapa
- Created unit tests and update/maintain documentation for Python packages supporting the company
- Utilized GitHub for software development, version tracking, and data analysis collaboration
- Utilized AWS Sagemaker and Cloudwatch to analyze/create ML models with data stored in S3
- Presented analytical findings to groups of 5-10 people in an understandable and concise manner
- Quickly switched between projects, adapting to different techniques in the start-up environment

Genetic Research Experience, Balciunas Lab, Temple University, PA *December 2018-May 2021*

- Effectively collaborated with a team to engineer a fully 'floxed' *tcf21* allele for the study of post-embryonic regenerative abilities of zebrafish hearts, and of the role *tcf21* plays in heart regeneration
- Created a pipeline for transforming large, raw ChIP-seq fastq datasets with GalaxyHub, bowtie 2, samtools, R and Python scripts, and MEME-ChIP with statistical programming practices
- Experienced with Gel Electrophoresis, PCR, Zebrafish Maintenance/Rearing, Survival Surgeries (*Danio rerio*), Conditional Mutagenesis, Cre-loxP, CRISPR/Cas9, Fluorescent Microscopy
- **Poster**: Conditional Mutagenesis of Zebrafish *tcf21* at 2019 Undergraduate Research Symposium
- Presented at weekly lab meetings with in-depth analysis and discussion of obtained data, as well as peer-reviewed journal presentations
- Accepted to Research Experiences for Undergraduates REU, CSUSM: Cancelled, COVID-19 Pandemic

Joshua P. Schaaf

610-550-6432 • Davis, CA • joshuaschaaf@gmail.com • https://github.com/viabard

PUBLICATIONS

Pang-ning Teng, **Joshua P. Schaaf**, et al. (2024). ProteoMixture: A cell type deconvolution tool for bulk tissue proteomic data. *iScience*, Volume 27 (3) https://doi.org/10.1016/j.isci.2024.109198

Allison L Hunt … **Joshua P. Schaaf**, et al. (2024). Mapping three-dimensional intratumor proteomic heterogeneity in uterine serous carcinoma by multiregion microsampling. *Clinical proteomics*, Volume 21 (1) https://doi.org/10.1186/s12014-024-09451-2

CODING PROJECTS

Imputation Analysis with High Dimensional Data: Using many techniques described in literature, including HoloClean in PostgreSQL, I analyzed different imputation methods for highly dimensional data.
Using Computer Vision ML for Species Recognition: Developed Python scripts to retrieve over 300,000 images of PA moths from the iNaturalist API, then a TensorFlow deep learning model for species recognition.
Using Machine Learning for Chest X-Ray Abnormality Detection: Created a machine learning model with Python scripts for CXR preprocessing with scikit-image and object detection with YOLOv5 in PyTorch.

EDUCATION

Temple University, College of Science and Technology — Philadelphia, PA
Professional Science Masters (PSM), Bioinformatics — *May 2022 - Cumulative GPA: 4.0/4.0*
Bachelor of Science, Biochemistry, Minor in Computer Science — *May 2021 - Cumulative GPA: 3.9/4.0*

Honors and Distinctions: ***Magna Cum Laude***, Distinction in Major, Dean's List, **Temple Honors Program**, **Natan Luehrmann-Cowen Memorial Award** (College-Wide Academic Excellence, Musical Instrument)
Scholarships: **NSF REU** at CSUSM with Dr. Arun Sethuraman (Cancelled due to SARS-CoV-2), **Science Scholars Program** (SSP), Temple Academic Scholarship, TUDMB Scholarship

RELEVANT COURSEWORK

Mathematics/CS:	Coding in C; Biological Models in Python; Program Design/Abstraction (Java); Data Structures (Java); Calculus I, II, Honors III; Machine Learning; Biostatistics
Physical Sciences:	Honors Organic Chemistry I, II; Physical Chemistry of Biomolecules; Physics I, II
Biological Sciences:	Genomics (R, Python); Genetics; Cell Structure and Function; Biochemistry I, II; Fundamentals of Genomic Evolutionary Medicine; Ethics in Biotechnology

LEADERSHIP AND ACTIVITIES

1st Flute, Video Game Orchestra at UC Davis — *2022*
Volunteer Belayer, Reach Climbing & Fitness — *2022*
Principle 2nd Violinist, Lower Merion Symphony — *2021-2022*
1st Flute/Piccolo Player, Temple University Diamond Marching Band/Collegiate Band — *2017- 2018*