

Lead Bioinformatics/Data Scientist

Accomplished Bioinformatics Scientist with robust background in machine learning, computational biology, and bioinformatics, specializing in application of advanced statistical models and AI-driven algorithms to genomic research and therapeutics.

Adept at leveraging cutting-edge technologies, including CRISPR, NGS, and deep learning frameworks, to develop innovative tools and pipelines for genetic analysis, with focus on streamlining complex processes and delivering impactful results. Proven track record in leading cross-functional project teams, developing high-performance algorithms, and optimizing data workflows to accelerate scientific discoveries. Demonstrated expertise in designing data-driven solutions for diverse biological challenges, from gene editing to pathogen detection, with emphasis on improving accuracy and enhancing decision-making. Strategic problem-solver passionate about mentoring aspiring scientists and advancing the field through continuous learning and collaborative innovation.

Core Competencies

- Computational Biology
- Genomic Data Analysis
- Advanced Statistical Modeling
- Machine Learning in Genomics
- Bioinformatics Algorithms
- AI & Deep Learning Applications
- Next-Generation Sequencing (NGS)
- Bioinformatics Software Development
- CRISPR & Gene Editing Technologies
- High-Performance Computing (HPC)
- Project & Team Leadership
- Biological Data Integration
- Genetic Data Interpretation
- Data Pipeline Optimization
- Research & Scientific Innovation

Professional Experience

Mammoth Biosciences, Brisbane, CA Senior Data Scientist, Bioinformatics

2021 to Present

Engineered and deployed scalable algorithms and pipelines for CRISPR guideRNA design, image recognition, and protein function prediction. Built intuitive dashboards with Streamlit and Redash, leveraging Dbt and AWS Athena for interactive dataset exploration. Designed multi-use statistical and machine learning models for omics data processing and protein engineering applications. Collaborated with cross-functional teams to standardize workflows and integrate computational tools for diagnostics and therapeutics. Conducted statistical analyses to interpret genomic and proteomic datasets, enabling actionable research insights. Created efficient ETL pipelines using regression models to support high-performing variant selection processes.

- Improved genetic target optimization, achieving 100,000+ guideRNA designs for CRISPR with AWS-deployed Prefect pipelines (DARPA-DIGET).
- Attained 98% accuracy in protein function predictions by developing deep-learning (CNN) algorithms on ESM embeddings.
- Increased image analysis throughput to 5,000+ monthly samples with reliable OpenCV workflows deployed in AWS.
- Enhanced team efficiency by 95% through in-house deployment of protein variant calling algorithm as AWS Batch Jobs.
- Secured 70% predictive accuracy for chemically modified gRNA performance via deployable NLP and breadth-first search pipeline.
- Reduced time-to-market by 40% for diagnostic tools by standardizing product development protocols.
- Boosted data processing efficiency by 25%, accelerating research decisions through optimized machine learning algorithms.
- Achieved 99% accuracy in COVID-19 variant identification with scalable statistical algorithm adaptable for other genotyping tasks.

Arc Bio LLC, Scotts Valley, CA Bioinformatics Analyst

2017 to 2021

Led Next-Generation Sequencing (NGS) initiatives for agnostic microbial identification in human samples. Improved cancer detection and microbe classification by optimizing TensorFlow models and Neural Networks. Delivered actionable insights through advanced statistical analysis of large omics datasets.

- Increased virus classification accuracy by developing Python pipeline using deep learning (CNN).

- Enhanced data processing speed by deploying high-memory instances with Dockerized NGS pipelines.
- Streamlined product launch processes by standardizing workflows, ensuring consistency.
- Strengthened AMR/AvR capabilities, improving predictive accuracy in analysis pipelines.

Affymetrix/Thermo Fisher Scientific, Santa Clara, CA**2014 to 2017****Bioinformatics Scientist/Engineer**

Led AgBio and Human Microarray design projects, enhancing genetic research capabilities. Developed tools for microarray data analysis, streamlining workflow and improving accuracy. Analyzed complex microarray data to extract actionable insights for genetic studies.

- Increased application versatility by designing and optimizing custom Axiom SNP arrays for agrigenomics and human genetics.
- Boosted data processing and reporting efficiency by automating analysis pipelines using Python and R.
- Enhanced Axiom platform microarray chip design, elevating performance and precision.

Education & Credentials

Professional Certificate in Machine Learning & AI, University of California, Berkeley | 2023

Master of Science, Bioinformatics, San Jose State University, San Jose, CA | 2014

Master of Philosophy, Neuro-Biophysics, National Institute of Mental Health & Neuro Sciences, India | 2010

Licenses & Certifications

NASA Open Science, NASA | October 2024

Professional Certificate in ML & AI, University of California, Berkeley | June 2024

Technical Proficiencies

- **Programming Languages & Software Development:** Python, R, Java
- **Bioinformatics Tools & Platforms:** BLAST, Clustal Omega, Bioconductor, Ensembl
- **Machine Learning & Artificial Intelligence:** Deep Learning Frameworks (e.g., TensorFlow, PyTorch), Data Modeling and Simulation, Neural Networks, Genetic Algorithms
- **Genomics & NGS Analysis:** Next-Generation Sequencing (NGS) Tools, CRISPR Design and Analysis, Variant Calling and Annotation, RNA-Seq and Transcriptomics
- **Database Management & Big Data:** SQL, NoSQL (MongoDB), Data Cleaning and Preparation, Pipeline Development
- **Visualization & Reporting:** Data Visualization Tools (Tableau, Power BI, R Shiny), Report Generation, Dashboard Design
- **Computational Biology:** Protein Structure Prediction, Molecular Dynamics Simulations, Systems Biology Models
- **High-Performance Computing:** Cluster Computing, Cloud Platforms (AWS, Azure), Parallel Processing

Publications / Patents

Peer-Reviewed Publications

- COVID-19 Variant Detection with a High-Fidelity CRISPR-Cas12 Enzyme. Journal of Clinical Microbiology, 60(7), e0026122. <https://doi.org/10.1128/jcm.00261-22>
- Metagenomic Next-Generation Sequencing for Identification and Quantitation of Transplant-Related DNA Viruses. Journal of Clinical Microbiology, 57(12): e01113-19. Published 2019 Nov 22. <https://doi.org/10.1128/JCM.01113-19> PMID: 31554674

Patents

- Programmable Nuclease Diagnostic Device, WO2022271873A3
- Systems and Methods for Identifying Genetic Phenotypes Using Programmable Nucleases, WO2023097325-A3
- Systems, Methods, and Media for Determining Relative Quality of Oligonucleotide Preparations, WO2022026905-A1
- Normalization Controls for Managing Low Sample Inputs in Next-Generation Sequencing, CA3115281A1