
SUMMARY

Performance-focused Machine Learning Scientist with a Ph.D. in Biological Sciences and more than eight years of experience architecting and deploying scalable AI/ML pipelines for genomics, clinical research, and multi-omics datasets. Proficient in cutting-edge neural network designs – including convolutional networks and transformer architectures – deep learning methods, and comprehensive MLOps workflows. A trusted leader and collaborator with biostatisticians, bioinformaticians, and clinicians, driving precision-oncology initiatives from concept through production. Passionately engaged with the latest advances in diffusion and LLM-driven protein-engineering applications.

CORE COMPETENCIES

- **AI/ML Frameworks & Algorithms:** Python ecosystem (NumPy, Pandas, scikit-learn, TensorFlow/Keras, PyTorch) including Keras Tuner; transfer learning and RAG; transformer architectures; and time-series forecasting.
- **Data Domains & Bioinformatics:** Designed and validated end-to-end bioinformatic pipelines (germline, somatic, RNA-seq, proteomics, metabolomics); feature selection; statistical analysis plan and inference; quality management in compliance with FDA IVD guidelines
- **MLOps:** Docker microservices and network management, MLflow, CI/CD pipelines; REST API for model sharing; Git/GitHub; model versioning, monitoring, and pruning.
- **Cloud & HPC Platforms:** AWS Batch, S3, and EC2; HPC job orchestration with SLURM; Nextflow workflow orchestration; and Docker container networking and deployment
- **Programming & Data Engineering:** Advanced skills in Python and R; SQL and PySpark for large-scale ETL and high-dimensional clinical data processing
- **Collaboration & Agile Delivery:** Agile methodologies with JIRA; cross-functional teamwork with biostatisticians, bioinformaticians, and clinicians; stakeholder expectation management

EXPERIENCE

DATA SCIENTIST III

Sapient Bioanalytics

Oct 2023 - present

- **Multi-Omics ML Pipelines:** Architected a three-stage analytical workflow for a 50,000-sample metabolomics study (40,000+ features/sample), integrating data cleaning, feature engineering (PySpark), and regression analysis.
- **Modeling:** Pioneered application of transformer architectures to longitudinal metabolomic data for risk-score prediction of preeclampsia, stillbirth, and preterm birth across five global cohorts.
- **Metabolic Risk Scoring:** Developed an XGBoost-driven scoring system with Optuna-optimized hyperparameters; deployed on Databricks to process population-scale datasets.
- **Capacity Building:** Designed and delivered training workshops in developing countries on best practices for multi-omics analysis and AI methodology, strengthening local bioinformatics capabilities.
- **Cross-Functional Collaboration:** Coordinated with mass spectrometry, computational chemistry, and clinical scientists to ensure data integrity, reproducibility, and alignment with project milestones.

BIOINFORMATIC SCIENTIST II - III

Ambry Genetics

Dec 2019 - Oct 2023

- **Panel Development:** Design and validation of high-volume NGS panels (CancerNext, RNA, WES, and somatic panels). Ensured $\geq 99\%$ sensitivity/specificity for SNVs, indels; CNVs under FDA regulatory guidelines.
- **QC & Statistical Thresholding:** Established robust statistical cutoffs for allele-frequency and coverage metrics, reducing false positives by 30% during chemistry transitions.

- **Pipeline Performance Monitoring:** Developed real-time dashboards to track and visualize variant-calling pipeline performance.
- **Machine Learning Applications:** Implemented GradientBoostingRegressor models to predict CNV counts and prioritize QC investigations, improving workflow efficiency.

Postdoctoral Researcher

UC Irvine

Jan 2014 - Dec 2019

- **CRISPR & Proteomics:** Developed CRISPR tagging/knockout tools and purified tagged PP2A complexes for SILAC-LC-MS; mapped protein-protein interactions under metabolic stress.
- **HTS Bio-Screening:** Developed high-throughput screening platform in mammalian cells for p53-reactivating compounds, resulting in a publication in *Nature Communications*.
- **RNA-seq Analytics:** Designed Nextflow pipelines for differential expression and splicing analysis, identifying methionine-stress targets for therapeutics.
- **Metabolomics Integration:** Processed and analyzed raw LC-MS metabolomic profiles; performed pathway enrichment to elucidate methionine addiction mechanisms in cancer.
- **The Data Incubator:** Completed a capstone project to visualize the LA parking **sweet spots**.

EDUCATION

University of California, Irvine

Sep 2007 - Jan 2014

Ph.D. Biological Sciences

University of California, Berkeley

M.S. Information and Data Science

In progress

AWARDS & PUBLICATIONS

- Molecular Metabolism** | Nutrient control of splice site selection contributes to methionine addiction of cancer 2025
- The EMBO Journal (Accepted)** | Sphingosine and anti-neoplastic sphingosine analogs activate PP2A and inhibit nuclear import in parallel by engaging PPP2R1A and importins 2025
- Cell Chemical Biology** | Discovery of compounds that reactivate p53 mutants in vitro and in vivo 2022
- Journal of Lipid Research** | Lipid remodeling in response to methionine stress in MDA-MBA-468 triple-negative breast cancer cells 2021
- Journal of Biological Chemistry** | Microhomology-based CRISPR tagging tools for protein tracking, purification, and depletion 2019
- Methods Mol. Biol.** | Isolation and characterization of methionine-independent clones from methionine-dependent cancer cells 2019
- U.S. Patent** | Chembridge Small Molecules that could enhance p53 activity 2015
- Journal of Cell Science** | SAM limitation induces p38 mitogen-activated protein kinase and triggers cell cycle arrest in G1 2014
- Nature Communications** | Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53 2013
- Cell Cycle** | Downregulation of Cdc6 and pre-replication complexes in response to methionine stress in breast cancer cells 2012
- Journal of Biological Chemistry** | Transforming growth factor β up-regulates cysteine-rich protein 2 in vascular smooth muscle cells via activating transcription factor 2 2008
- Genes to Cells** | Identification of a putative human mitochondrial thymidine monophosphate kinase associated with monocytic/macrophage terminal differentiation 2008