Niu Du, PhD

Phone: +1 (858) 603-5663 | **Email:** ndu0328@gmail.com | **GitHub: https://github.com/ndu-bioinfo**

Areas of Expertise

Bioinformatics & Computational Biology

Multiomics data analysis and integration of computational methods with biological insights.

Machine Learning & AI for Biological Data

Application of machine learning and AI models for analyzing complex biological datasets, such as genomics, drug discovery, and bioinformatics pipelines.

Pipeline Development & Engineering

Design and implementation of scalable, high-throughput pipelines for data processing, quality control, and automation.

Data Management & Integration

Proficient in managing, integrating, and maintaining large datasets, including database management, API development, and efficient data retrieval.

Technical Skills

Programming Languages

Python, R, SQL, Bash; scikit-learn, biopython, plotly, PyTorch, FastAPI, Flask, Streamlit

Cloud & HPC Computing

SGE, GCP, AWS, Docker, Kubernetes, Snakemake, CI/CD, API development

Bioinformatics

NGS pipelines, RNA-seq, DNA-seq, single-cell, CITE-seq, 16S/18S, metagenomics, metatranscriptomics, variant calling, oligo design, microbiome analysis, deep learning, machine learning, Bayesian statistics, IVD assay development, quality control, sequencing platform optimization

Professional Experience

Visiting Bioinformatics Scientist

J. Craig Venter Institute

May 2024 – Current

- Researched and developed an optimized workflow for Bayesian inference to select the best model for fitting undersampled biological data, utilizing a coupled approach of simulated annealing and Markov Chain Monte Carlo (MCMC).

- Built a tools package to implement this workflow, enabling efficient application of Bayesian inference, and tested its performance on real-world pharmacological and clinical datasets.

Senior Bioinformatics Scientist

Creyon Bio

September 2022 – May 2024

- Developed and implemented next-generation sequencing (NGS) pipelines for QuantSeq and allele-specific expression (ASE) analysis, enhancing the accuracy of gene expression studies.
- Designed sparse grid Bayesian optimization for target detection in antisense oligonucleotide (ASO) drug development.
- Led the streamlining of bioinformatics workflows, ensuring smooth integration of data from research teams across production and R&D for increased efficiency.
- Integrated NGS pipelines with cloud platforms, developing APIs to ensure seamless data flow between storage systems, databases, and visualization dashboards, facilitating more effective data science workflows.
- Performed extensive data analyses to optimize pipeline performance, refining quality control processes to enhance the accuracy and efficiency of target identification in drug discovery.
- Established design and technical guidelines for pipeline frameworks, standardizing implementation and operational best practices in bioinformatics and drug discovery projects.

Bioinformatics Engineer, Algorithm

Invitae Corporation

June 2020 – September 2022

- Led the development of in vitro diagnostic (IVD) targeted DNA diagnostic assays for clinical use, contributing to successful product launches.
- Developed an automated oligo design workflow, reducing sequencing costs by 45% and increasing operational efficiency.
- Developed machine-learning-based target classification algorithms to improve assay specificity and sensitivity.
- Directed the development of bioinformatics tools from concept to library integration, ensuring robustness and reliability through comprehensive testing.
- Troubleshot pipeline operational issues and provided emergency response to address critical issues for customers, ensuring system stability and reliability.
- Actively participated in code reviews, planning sessions, and best practices discussions, fostering a collaborative and high-quality development environment.

Postdoctoral Fellow

La Jolla Institute for Immunology

April 2019 – June 2020

- Developed RNA-Seq mapping and QC pipelines for bulk and single-cell sequencing data.

- Created CITE-seq pipelines for single-cell data preprocessing and clustering, improving data analysis workflows.
- Collaborated with experimentalists on bulk and single-cell data analysis related to lung diseases.

Postdoctoral Fellow

J. Craig Venter Institute
April 2018 – April 2019

- Developed analytical pipelines for 16S, 18S, and metatranscriptome data, as well as transcription factor-DNA-binding assays.
- Conducted statistical analyses on human oral microbiome data for caries detection using mixed-effect models.

Research Associate

J. Craig Venter Institute
2015 – 2017

- Investigated metabolic pathways in photosynthesis using genome-scale models, contributing to a deeper understanding of energy utilization.
- Developed a novel method for rapid carbon fixation rate measurement, including system integration, protocol development, and software creation.
- Created a deep learning model for two-factor RNA-seq data analysis, enhancing data interpretation capabilities.
- Performed various wet lab experiments and routine bioinformatics tasks, including pipeline execution and server maintenance.

Professional Trainings and Certificates

- Microbial Genomics & Metagenomics, Joint Genome Institute, 2018
- Bioinformatics Genome Assembly, UC Davis, 2018
- IBM Data Science Professional Certificate, 2018
- Neural Networks and Deep Learning Certificate, 2017
- Micro MBA, Rady School of Business, 2016
- Innovation to Market, Rady School of Business, 2015

Education

Ph.D. in Biology
University of California, San Diego Scripps Institution of Oceanography 2011 – 2018
Research focus: Application of systems biology and bioinformatics on modeling marine microorganism pathway kinetics

B.S. in Applied Chemistry
Ocean University of China, Qing Dao 2002 – 2006

Publications

- **P1**: Broddrick, J. T., Du N. (Co-1st author), et al. (2019). Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in diatom Phaeodactylum tricornutum. New Phytologist.
- **P2**: Du N., et al. (2018). Simultaneous quantum yield measurements of carbon uptake and oxygen evolution in microalgal cultures. PLoS One.
- **P3**: Broddrick, J. T., et al. (2016). Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. PNAS, 113(51), E8344-E8353.
- **P4**: Woertz, I. C., et al. (2014). Life cycle GHG emissions from microalgal biodiesel–a CA-GREET model. Environmental Science & Technology, 48(11), 6060-6068.