

Niu Du, PhD

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

Areas of Expertise

Bioinformatics & Computational Biology

Multomics data analysis, integrating computational techniques with biological insights to drive advancements in genomics, transcriptomics, and other omics fields.

Machine Learning & AI for Multi-omics Data Analysis

Applied machine learning models to extract meaningful insights from complex biological datasets, including genomics, microbiome, and single-cell data, supporting precision medicine and drug development.

Pipeline Development & Engineering

Designed and implemented scalable, high-throughput pipelines for data processing, quality control, and automation, ensuring reproducibility and efficiency in bioinformatics workflows.

Data Management & Integration

Proficient in managing and integrating large-scale datasets, with expertise in database management, API development, and efficient data retrieval to support seamless data analysis and visualization.

Technical Skills

Programming Languages

Python, R, SQL, Bash

Cloud & HPC Computing

SGE, GCP, AWS, Docker, Kubernetes, 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

- Designed and implemented a Bayesian inference workflow optimized for non-linear model optimization, significantly improving the detection of targets in dose-response data. This approach uncovered 976 (20.3%) additional targets in chemotherapy drug discovery (paper in preparation).

- Developed a generalized tool package for the Bayesian workflow, extending its application across diverse biological datasets, including enzyme activity and tumor growth data. This toolkit has laid the groundwork for applying probabilistic modeling in other omics data types, facilitating broader adoption of Bayesian methods in biological research.

Senior Bioinformatics Scientist

Creyon Bio

September 2022 – May 2024

- Developed and implemented next-generation sequencing (NGS) pipelines for QuantSeq and allele-specific expression (ASE) assays, establishing production-stable pipelines from scratch. These pipelines formed the foundation for Creyon Bio's successful efforts in TNPO2 drug discovery, contributing to the company's advancements in rare disease therapy[<https://globalgenes.org/raredaily/644745/>].
- Designed a sparse grid Bayesian optimization algorithm for target detection in antisense oligonucleotide (ASO) drug development, improving the identification of therapeutic targets and optimizing experimental designs for efficient drug screening.
- Applied machine learning models, including neural networks and XGBoost, to enhance toxicity prediction, integrating predictive analytics into the drug development process and improving the selection of viable drug candidates.
- Led workflow streamlining efforts, improving cross-departmental data integration between production and R&D teams, thereby increasing the overall efficiency of the bioinformatics pipeline and reducing data processing times.
- Integrated cloud-based solutions, creating APIs that facilitated seamless data flow between cloud storage systems, databases, and visualization platforms. This integration enabled real-time data analytics and fostered a more data-driven approach to drug discovery.
- Performed in-depth data analyses to optimize pipeline performance, refining quality control processes to improve the reliability of NGS data and enhance the accuracy of target identification in drug discovery projects.
- Established technical guidelines for pipeline frameworks, ensuring consistency in bioinformatics operations and providing a foundation for the scalable development of future bioinformatics workflows

Bioinformatics Engineer, Algorithm

Invitae Corporation

June 2020 – September 2022

- Led the development of in vitro diagnostic (IVD) targeted DNA diagnostic assays for clinical use, directly contributing to the successful launch of multiple diagnostic products used in clinical settings, ensuring high sensitivity and specificity in detecting genetic variants.
- Developed an automated oligo design workflow that significantly reduced sequencing costs by 45%, optimizing reagent use and streamlining lab operations for greater operational efficiency across multiple diagnostic and carrier screening assay development projects.
- Implemented machine-learning-based target classification algorithms to streamline and systematically improve the efficiency of the QA process for called genetic variants. This innovation reduced the time

required for assay quality assessment from two months to two weeks, accelerating the validation workflow and improving the overall throughput of genetic data analysis.

- Directed the end-to-end development of bioinformatics tools, from initial concept to full integration into production libraries, ensuring robustness and reliability through rigorous performance testing and validation processes.
- Managed troubleshooting and provided emergency support to resolve critical pipeline issues, maintaining system uptime and stability for customers, and ensuring continuous diagnostic services in clinical workflows.
- Actively contributed to code reviews, planning sessions, and best practices discussions, fostering a collaborative development environment focused on quality and continuous improvement in bioinformatics tool development.

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.