

Job Title: AI Scientist (LLM for Multi-Omics & Precision Medicine)

About the Lab

The Dong Lab (<http://www.donglab.org>) at Yale School of Medicine is seeking a highly motivated AI-focused scientist to develop next-generation large language model (LLM) approaches for gene regulation and precision medicine.

Our lab leads the **NEUROMICS platform**, a large-scale effort integrating:

- Thousands of human brain samples with single-cell and bulk multi-omics
- Multi-layer regulatory features (splicing, polyadenylation, circRNA, RNA modification, eRNAs, xQTLs, etc.)
- Millions of longitudinal EHR records (e.g., COSMOS and other real-world datasets)

We have built a robust data foundation and standardized pipelines. The next phase is to leverage AI/LLMs to transform these datasets into predictive and mechanistic models of human disease.

Position Overview

We are looking for a creative and driven researcher to develop and apply LLM-based models to large-scale biological and clinical datasets.

This role focuses on building **foundation models for gene regulation and disease**, with opportunities to connect computational predictions to **organoid-based experimental validation**, forming a closed-loop discovery system.

Key Responsibilities

You will lead or contribute to one or more of the following directions:

1. Virtual Cell Modeling

- Develop LLM-based or foundation models (e.g., extending GeneFormer-like architectures)
- Model gene regulatory programs across cell types and conditions
- Integrate multi-modal features beyond gene expression (e.g., splicing, RNA processing, regulatory elements)

2. Disease Modeling

- Build predictive models for neurodegenerative diseases (PD, ALS, AD)
- Integrate multi-cohort omics data to model disease onset, progression, and heterogeneity
- Link molecular states to clinical phenotypes using large-scale EHR data

3. Therapeutic Discovery

- Perform AI-driven drug repurposing using real-world data
- Infer therapeutic targets via gene–drug interaction resources (e.g., CMAP-like datasets)
- Develop models for drug–compound similarity and novel therapeutic inference

4. Model Development & Engineering

- Fine-tune or pretrain large-scale models on multi-omics datasets
- Design scalable pipelines for training and evaluation
- Contribute to open-source tools and reproducible workflows

Required Qualifications

- Strong coding skills (Python required; experience with ML frameworks such as PyTorch, JAX, or TensorFlow)
- Demonstrated experience or strong interest in AI/ML modeling (especially deep learning or LLMs)
- Ability to work with large-scale datasets (cloud/HPC experience preferred)
- At least one first-author publication or major project (paper, preprint, or equivalent work)
- Strong problem-solving skills and ability to work independently

Preferred Qualifications (not required)

- Experience with LLMs or foundation models (e.g., transformers, generative models)
- Exposure to computational biology, genomics, or biomedical data
- Familiarity with gene regulation concepts
- Experience with sequencing data analysis
- Active GitHub contributions or open-source involvement
- Experience with AI-assisted coding tools (e.g., Claude Code, Copilot)
- Background in multi-modal data integration

Who Should Apply

We welcome candidates from diverse backgrounds, including:

- Computer science / AI / data science
- Physics / math / engineering
- Computational biology / bioinformatics

A formal background in molecular biology is **not required**, but curiosity about biological systems is important.

What We Offer

- Access to **unique, large-scale multi-omics and EHR datasets**
- Strong computational infrastructure (GPU clusters, cloud resources)
- Close integration with experimental platforms (organoids, in vivo models)
- Highly collaborative environment across Yale and external partners
- Opportunity to lead high-impact projects at the intersection of AI and medicine

Application

Please send:

- CV
- Brief statement of research interests
- GitHub (if available)
- Representative work (papers, preprints, or projects)

to: xianjun.dong@yale.edu