

# Kaeli Rizzo

---

## CONTACT INFORMATION

+1-973-464-5228  
rizzo@cshl.edu  
kaeliarizzo

kaeli-rizzo-5575ba1b6  
orcid.org/0000-0001-5731-2046

## RESEARCH INTERESTS

Developing **efficient foundation models** through distillation and architectural optimization that enable **scalable and accessible** AI for scientific discovery.

## TECHNICAL SKILLS

Experienced with model development, finetuning, and interpretation (attribution analysis, in-silico mutagenesis) using distributed/parallel computing on HPC systems.

- **Programming Languages:** Python, Shell scripting, Java, MATLAB, and C.
- **ML & Data Science:** PyTorch, TensorFlow, Weights & Biases, Pandas, NumPy, SciPy, Matplotlib, Jupyter.
- **Computing Environment:** Linux, Git/GitHub, HPC clusters (UGE, Slurm), Bash, Conda, uv.
- **Bioinformatics Tools:** Minimap2, Samtools, Bwa, and Busco.

## EDUCATION

### Cold Spring Harbor Laboratory - Simons Center for Quantitative Biology, NY

**Ph.D. Candidate - School of Biological Sciences** 2023 – 2027 (expected)

- Thesis Topic: *Enabling Large-Scale Interpretation of Genomic Foundation Models through Knowledge Distillation*
- Adviser: Professor Peter K. Koo

### Georgia Institute of Technology, May 2022

- B.S., Biomedical Engineering, *Summa cum Laude*
- Minor in Computer Science (Computing and Intelligence)

## EXPERIENCE

### Koo Lab at Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

**Ph.D. Candidate**

August 2023 - Current

- Developed resource-efficient ensemble distillation framework that quantifies uncertainty from 1-3 teacher queries, bypassing full ensemble deployment.
- Optimized Enformer through FlashAttention and architecture reduction, achieving 4-5x faster inference and 2x memory efficiency while maintaining accuracy.
- Finetuned Enformer, a genomic foundation model, to model how aging impacts breast cell responses to pregnancy hormones
- Built multitask framework for chromatin accessibility prediction that quantifies both epistemic and aleatoric uncertainty and improves robustness under distribution shifts.

### Ma'ayan Lab - Icahn School of Medicine at Mount Sinai, New York, NY

**Bioinformatics Consultant**

December 2021 - March 2022

- Role required collaborating on cross-functional projects in bioinformatics.

**Summer Fellow**

June - August 2019

- Developed an SVM model to predict understudied protein-phenotype associations based on their relationship with well-known proteins.

### Jordan Laboratory - Georgia Institute of Technology, Atlanta, GA

**Undergraduate Research Assistant**

August 2020 - December 2021

- Constructed a socioeconomic deprivation measurement using PCA analysis on UK Biobank, improving upon traditional Townsend index for health disparity research.
- Performed comprehensive data cleaning and outcome modeling using UK Biobank data to identify key environmental drivers of ethnic health disparities.

## Khatri Lab – Stanford, Stanford, CA

### Summer Undergraduate Research Assistant

June - July 2021

- Analyzed scRNA sequencing data in Python on a High-Performance Computer (HPC).

## Regeneron Bioinformatics Core Services, Tarrytown, NY

### Co-op

January - August 2020

- Engineered a reproducible *de novo* genome assembly pipeline using PacBio and Illumina data, with automated installation scripts and comprehensive documentation for reproducible HPC deployment.

## Machine Learning Group - Georgia Institute of Technology, Atlanta, GA

### Undergraduate Research Assistant

August - December 2019

- Developed machine learning algorithms to analyze relationships between nanoparticle properties (lipid chain length, composition, size) and their functional effectiveness.

## JOURNAL PUBLICATIONS

- [1] **Kaeli Rizzo** and Peter K. Koo. Ensemble Distillation with Stochastic Teachers via Online Moment Estimation. In Progress.
- [2] Jessica Zhou, **Kaeli Rizzo**, Ziqi Tang, Peter K. Koo. *Uncertainty-aware genomic deep learning with knowledge distillation*. bioRxiv (2024).
- [3] Alexander Lachmann, **Kaeli Rizzo**, Alon Bartal, Minji Jeon, Daniel J. B. Clarke, Avi Ma'ayan. *PrismEXP: gene annotation prediction from stratified gene-gene co-expression matrices*. PeerJ (2023).

## SELECTED TALKS

- *Enabling Large-Scale Interpretation of Genomic Foundation Models*. Quantitative Biology and Artificial Intelligence (QB-AI) Seminar. CSHL, NY. March, 2025.
- *Enabling Large-Scale Interpretation of Genomic Foundation Models through Knowledge Distillation*. Graduate Student Symposium. CSHL, NY. December, 2024.

## POSTERS

- *Efficient Knowledge Distillation for Genomic Deep Learning*. Keystone Symposium - AI in Molecular Biology. Sante Fe, NM. September, 2025.
- *Enabling Large-Scale Interpretation of Genomic Foundation Models*. CSHL Biology of Genomes. Cold Spring Harbor, NY. May, 2025.
- *Mammalian De Novo Whole Genome Assembly*. Regeneron Symposium. Tarrytown, NY. August, 2020.

## MENTORSHIP

### Summer Research Mentor

2025

- Mentored high school student on project using surrogate models to interpret gene regulatory elements from genomic foundation models.

### NSF Postbaccalaureate Research Education Program Mentor

2024-2025

- Mentored 2 NSF postbaccalaureate researchers through graduate school applications: school selection, application strategy, and NSF GRFP proposal development.

## PROFESSIONAL SERVICE

### Referee Service

- Research in Computational Molecular Biology (RECOMB) (2025)
- Nature Methods (2024, 2025)