

Kaeli Rizzo

CONTACT INFORMATION

✉ rizzo@cshl.edu
📧 kaeliarizzo

🔗 kaeli-rizzo-5575ba1b6
🏠 kaeliarizzo.github.io
orcid.org/0000-0001-5731-2046

RESEARCH INTERESTS

Developing reliable biological sequence models by quantifying uncertainty, testing generalization to new sequences, and identifying what data and modeling strategies would improve prediction reliability.

TECHNICAL SKILLS

Experience developing, fine-tuning, and interpreting biological sequence models, with distributed training across clusters and cloud infrastructure.

- **Programming Languages:** Python, Shell scripting, Java, MATLAB, and C.
- **ML & Data Science:** PyTorch, TensorFlow, Weights & Biases, Pandas, NumPy, SciPy, Matplotlib, Jupyter.
- **Systems & Compute:** Linux, Git/GitHub, SLURM, UGE, AWS, Run:AI, 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 2022 – January 2027 (expected)

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

Georgia Institute of Technology, May 2022

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

RELEVANT 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 a single teacher query, eliminating the need for full ensemble access. Evaluated predictive performance and uncertainty calibration on genomic sequence-to-function and image classification benchmarks.
- Engineered compact, inference-efficient version of Enformer, a genomic foundation model, through FlashAttention and architectural optimization. Achieved 4-5x faster inference and 2x memory efficiency while maintaining accuracy.
- Built multitask framework for chromatin accessibility prediction that quantifies epistemic and aleatoric uncertainty, with improved robustness under distribution shifts.
- Fine-tuned and interpreted Enformer to model age-related alterations in cis-regulatory mechanisms that regulate breast tissue response to pregnancy hormones.

Sanofi, Cambridge, MA

RNA Machine Learning Research Co-op

July 2026 - Current

- Developed supervised and self-supervised RNA sequence models and refined benchmarks to evaluate downstream predictive performance.
- Delivered a team-wide presentation to 30+ colleagues on effective, cost-conscious use of agentic coding tools, covering context management and workflow best practices.

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.

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

Summer Fellow

June - August 2019

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

PUBLICATIONS & MANUSCRIPTS

- [1] **Kaeli Rizzo**, Yiyang Yu, and Peter K. Koo. Teacher-Access-Constrained Ensemble Distribution Distillation. Manuscript.
- [2] **Kaeli Rizzo***, Caroline M. Weisman*, Richard Sever, and Peter K. Koo. The scientific paper needs an uncertainty layer. Preprint.
- [3] Masayuki Nagai, Alan E. Murphy, **Kaeli Rizzo**, and Peter K. Koo. Toward generalizable and interpretable AI in regulatory genomics. *Nature Genetics* (2026).
- [4] Jessica Zhou, **Kaeli Rizzo**, Trevor Christensen, Ziqi Tang, and Peter K. Koo. Uncertainty-aware genomic deep learning with knowledge distillation. *npj Artificial Intelligence* (2026).

*Equal contribution.

TALKS

- *Enabling Large-Scale Interpretation of Genomic Foundation Models*. Quantitative Biology and Artificial Intelligence (QB-AI) Seminar. CSHL, NY. March, 2025.

POSTERS

- *Efficient Knowledge Distillation for Genomic Deep Learning*. Selected spotlight poster, MLCB 2026. New York, NY. November 2026.
- *Enabling Large-Scale Interpretation of Genomic Foundation Models*. CSHL Biology of Genomes. Cold Spring Harbor, NY. May, 2025.

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.

OUTREACH

Graduate School Application Workshop Volunteer

2023, 2025

- Reviewed personal statements and provided application guidance to prospective graduate students through CSHL's DIAS workshop.

WiSE Youth Coding Camp Instructor

2024

- Developed curriculum and taught Python programming to high school students through CSHL's Women in Science & Engineering (WiSE) program.

PROFESSIONAL SERVICE

Referee Service

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