Kaeli Rizzo

Contact Information

rizzo@cshl.edu
kaeliarizzo
kaeli-rizzo-5575ba1b6
kaeliarizzo.github.io
orcid.org/0000-0001-5731-2046

Research Interests

Developing **efficient foundation models** through distillation and architectural optimization that enable **scalable and accessible**.

Technical Skills

Model development, finetuning, and interpretation (attribution analysis, counterfactual prediction) using distributed training on High Performance Computing clusters.

- **Programming Languages**: Python, Shell scripting, Java, MATLAB, and C.
- **ML & Data Science**: PyTorch, TensorFlow, Weights & Biases, Pandas, NumPy, SciPy, Matplotlib, Jupyter.
- **Systems**: Linux, Git/GitHub, HPC clusters (SLURM, UGE), 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: Dr. 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 a single teacher query, eliminating the need for full ensemble access.
- 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.
- Finetuned and interpreted Enformer to model age-related alterations in cis-regulatory mechanisms that regulate breast tissue response to pregnancy hormones.

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.

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.

Journal Publications

[1] **Kaeli Rizzo** and Peter K. Koo. *Ensemble Distillation with Stochastic Teachers via Online Moment Estimation.* In Preparation.

[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. Santa 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.

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)