

TAYKHOO DALAL

tid4007@med.cornell.edu | <https://taykhoomdalal.github.io> | <https://www.linkedin.com/in/taykhoom-dalal>

EDUCATION

Tri-Institutional PhD Program (Cornell + MSK + Rockefeller) *Aug 2023 - Present*
PhD student in Computational Biology and Medicine 4.30 GPA

University of California, Los Angeles (UCLA) *Sept 2019 - June 2023*
Bachelor of Science (B.S.) in Computer Science Summa Cum Laude: 3.98 GPA

EXPERIENCE

Microsoft Research, Boston, MA
BioML Research Intern *June 2026 – Present*

- Developing genomic knowledge distillation approach for distilling privileged information from MSA-based “teacher” model to single-sequence model for inference time flexibility

Memorial Sloan Kettering, New York City, NY
PhD Student - Dr. Quaid Morris Lab *September 2024 - Present*

- Published SOTA mRNA foundation model called Orthrus, outperforming existing models with 700x more parameters (Orthrus 10M vs Evo2 7B) on RNA property and function prediction

Weill Cornell Medicine / Memorial Sloan Kettering, New York City, NY
PhD Rotation Student *Aug 2023 - August 2024*

- *Quaid Morris Lab:* Benchmarked RoseTTAFold2NA and AlphaFold 3 for nucleic acid-protein complexes folding, showing finetuned ESM-2 models achieve superior performance
- *Christina Leslie Lab:* Benchmarked DNA sequence to function models for cell-type-specific chromatin accessibility prediction
- *Fei Wang Lab:* Collaborated with Intelligent Data Analytics Lab @ Michigan State University to explore how LLMs could be used to improve Alzheimer’s disease risk prediction on longitudinal EHR data

Harvard Medical School, Boston, MA
Research Intern - Dr. Chirag Patel Lab *June 2022 - Feb 2024*

- Designed a high-throughput pipeline for phenome-wide association studies (PheWAS) on UKBB data
- Built a Python tool for PheWAS execution, visualization, and gene/variant annotation via REST APIs

Memorial Sloan Kettering Cancer Center, New York City, NY
Research Intern - Dr. Michael Berger Lab *June 2021 - Sept 2021*

- Built a workflow to clean and preprocess large-scale MSK genomic data, annotate variants via parallel database queries, and predict germline missense pathogenicity using random forest model

Northrop Grumman, Remote

Software Engineering Intern

June 2020 - Aug 2020

- Interned in the Algorithmic Warfare and Intelligence Department, working on a defense software for the U.S. Navy
- Developed GUI features, removed situation-dependent exceptions, and reduced memory load by up to 80% using a multiple ring buffer approach

PUBLICATIONS

1. Shi R.*, **Dalal T.***, Fradkin P.*, et al. 2025. mRNA-Bench: A curated benchmark for mature mRNA property and function prediction. *In Review*. <https://doi.org/10.1101/2025.07.05.662870>
 - a. *Spotlight at Machine Learning in Computational Biology (MLCB) 2025*
 - b. *Poster at AI in Molecular Biology 202, ICLR 2026 Workshops (LMRL, MLGenX)*
2. Fradkin P.*, Shi R.*, **Dalal T.***, et al. 2025. Orthrus: Towards evolutionary and functional RNA foundation models. *Nature Methods*. <https://www.nature.com/articles/s41592-026-03064-3>
 - a. *Oral at RNA-AI Conference 2025*
3. Sasse A., Ray D., Lavery K. U., Tam C. L., Albu M., Zheng H., Lyudovik O., **Dalal T.**, et al. 2025. A resource of RNA-binding protein motifs across eukaryotes reveals evolutionary dynamics and gene-regulatory function. *Nature Biotechnology*. <https://www.nature.com/articles/s41587-025-02733-6>
4. Mehine M., Muldoon D., Perry M. A., Kemel Y., Fong C., Gormally M.V., Jeng M.Y., Ge J., Tischfield S.E., Safonov A., Waters M., Pichotta K., Arora K., Syed A., Shah R., **Dalal T.**, et al. 2025. Somatic zygosity of germline BRCA1/2 mutations dictates lineage-agnostic HRD-phenotypes and platinum-sensitivity. *In Submission*.
5. Wang J., Ahn S., **Dalal T.**, et al. 2024. Augmented Risk Prediction for the Onset of Alzheimer's Disease from Electronic Health Records with Large Language Models. *arXiv*. <https://arxiv.org/abs/2405.16413>.
 - a. *Oral at American Medical Informatics Association NLP (AMIA NLP) 2025*
6. Parada G., Bretschneider H., Li J., Fine J., Dupas S., Ellis J., Braunschweig U., Bonnal S., **Dalal T.**, et al. 2024. An expanded repertoire of brain microexons is directly impacted by autism-linked genetic variation. *Nature (In Press)*.
7. **Dalal T.** and Patel C. 2024. PYPE: A pipeline for phenome-wide association and Mendelian randomization in investigator-driven biobank scale analysis. *Cell Patterns*. <https://doi.org/10.1016/j.patter.2024.100982>
8. Lee D., Donald K. A., **Dalal T.**, et al. 2020. Brain Network Connectivity from matching cortical feature densities. *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. <https://doi.org/10.1109/ISBI45749.2020.9098689>

* Co-first authorship (Equal Contribution)

AWARDS AND SCHOLARSHIPS

NSF Graduate Research Fellowship (2023)	NSF - 2023
UCLA Faculty Women's Club Endowed Scholarship (2023)	UCLA - 2023
Youngsoo Cha Endowed Scholarship in Engineering (2023)	UCLA - 2023
Kenneth I. Friedman Endowed Scholarship in Engineering (2021)	UCLA - 2021
Smita Patel Scholarship in Engineering (2020)	UCLA - 2020

TECHNICAL SKILLS

Programming: Python, Shell, C/C++, R

ML: PyTorch (Lightning; multi-GPU DDP/FSDP, AMP), Hugging Face, scikit-learn

BioML Benchmarking: Structural modeling (AlphaFold 3, RoseTTAFold2NA), Sequence models (Evo2, HyenaDNA, Nucleotide Transformer, Orthrus, etc)

Infra/Data: NumPy, pandas, SciPy, PyArrow/Parquet, conda/mamba, SLURM/LSF, Git/GitHub, Weights & Biases, Jupyter

LEADERSHIP AND SERVICE

Computational Biology Summer Program

Memorial Sloan Kettering

Journal Club Lead

June 2024 - August 2025

- Planned, set up, and led interactive journal clubs for the undergraduate computational biology summer interns for the CBSP program at MSK during the summer, redesigning the journal club experience to be more collaborative and interactive

Upsilon Pi Epsilon Honor Society

UCLA

Computer Science Tutor

Sept 2020 - June 2023

Entrepreneurship Chair

April 2022 - June 2023

- Hosted 2-4 tutoring hours every week for undergraduate C.S. students
- Planned and set up events focused on increasing the representation of students from diverse backgrounds in technology, inviting underrepresented individuals in the technology industry to come to UCLA to speak about their roads to success as entrepreneurs
- Set up events for introductory events with entrepreneurs from specialized fields in computer science, such as computational climate science and psychology, to help broaden the heterogeneity of career paths for computer science undergraduates at UCLA