AAKARSH ANAND

aakarsh@cs.ucla.edu · aakarsh-anand.github.io
Google Scholar · GitHub · LinkedIn

Machine learning researcher working on foundation models and statistical genetics for health data

EDUCATION

University of California, Los Angeles (UCLA) — Los Angeles, CA
Ph.D., Computer Science · Advisor: Sriram Sankararaman — 2024 – 2029 (expected)
Major field: Artificial Intelligence · Minor fields: Data Science Computing, Computational Systems Biology
M.S., Computer Science (earned during the Ph.D.) — 2026
B.S., Computer Science — 2020 – 2024

RESEARCH EXPERIENCE

Graduate Student Researcher, Sankararaman Lab, UCLA (undergraduate researcher 2022–24) — Jan 2022 – present

Foundation models for wearable sensors · with Yuzhe Yang's lab, UCLA

- One of three co-first authors of **Inertia-1** (NeurIPS 2026). Pretrained motion foundation models from 5M to 100M parameters, using recurrent, Transformer, ViT, and wav2vec-style architectures, on **18.2M hours** of accelerometry from **115K+ people** in NHANES and UK Biobank. Tested how sensor type, body placement, sampling rate, window length, architecture, model size, pretraining objective, and data scale affect performance on 15 downstream datasets covering activity recognition, freezing of gait, and disease prediction.
- Designed and ran the sensor-fusion study, which found that combining body placements and sensor types consistently beats single-sensor models.
- Co-first author of the follow-up journal paper (in prep.), which applies the model to diagnosis, prognosis, quantitative traits, and disease progression in five cohorts. Three of them (NHATS, ALS, and the Personalized Parkinson Project) are held out from pretraining. The paper also studies the genetics of the learned embeddings and early ways of pairing them with LLMs.

Deep learning for genetics

- Developing a training loss that makes a network's learned trait heritable and genetically correlated with disease, so a GWAS on it can find loci the original measurement misses. The genetic terms are computed exactly inside the training loop, and the loss works with any architecture.

Statistical genetics at biobank scale

- Leading a genome-wide test for gene–gene interactions in rare disease. It extends FAME to carriers of pathogenic variants to study how genetic background changes penetrance. Supported by an Anthropic AI for Science grant.
- Leading a single-cell variance-component model that separates cell-type-shared and cell-type-specific genetic and environmental effects on gene expression, and follows them along continuous cell states.
- Co-leading an atlas of genetic correlation across 3,570 UK Biobank trait pairs, partitioned by tissue-specific gene expression in 53 tissues.
- Ran simulations (calibration, power, runtime) for QuadKAST, FAME, and ENGINE and real-data analyses for FLEX, and added FAME's LD-block removal step.

Research Intern (NSF REU), Bruins in Genomics Summer Program, UCLA — Jun – Sep 2022

- Used symbolic regression to explain gene–gene interactions learned by nonlinear models (first-author poster).

Research Assistant, Boutros Lab, UCLA — Jan 2021 – Apr 2022

- Added the HISAT2 aligner to the lab's Nextflow alignment pipeline, redesigned its configuration, and benchmarked it. The pipeline is now part of Metapipeline-DNA (Cell Reports Methods 2026).

PUBLICATIONS

* equal contribution

Peer-reviewed

- Inertia-1: An Open Exploration of Wearable Motion Foundation Models
  Z. Xu*, **A. Anand***, S. Jiang*, C. Zhuang, Z. Shuai, S. Sankararaman, Y. Yang · **NeurIPS 2026** · code
- A biobank-scale method for learning modulators of gene-environment interaction underlying human complex traits from multiple environmental exposures
  Z. Liu, A. Ramteke, **A. Anand**, A. Gorla, M. Jeong, S. Sankararaman · **Genome Research** 2026 · selected from RECOMB 2026
- Metapipeline-DNA: A comprehensive germline and somatic genomics Nextflow pipeline
  Y. Patel, C. Zhu, T. N. Yamaguchi, N. K. Wang, ..., **A. Anand**, ..., P. C. Boutros (43 authors) · **Cell Reports Methods** 2026

- A biobank-scale test of marginal epistasis reveals genome-wide signals of polygenic interaction effects
  B. Fu, A. Pazokitoroudi, Z. Shi, A. Kar, A. Xue, **A. Anand**, P. Anand, Z. Liu, R. Border, P. Pajukanta, N. Zaitlen, S. Sankararaman · **Nature Genetics** 57(12):3175–3184, 2025
- A scalable adaptive quadratic kernel method for interpretable epistasis analysis in complex traits
  B. Fu*, P. Anand*, **A. Anand***, J. Mefford, S. Sankararaman · **Genome Research** 34(9):1294–1303, 2024 · RECOMB 2024

Under review

- Comprehensive gene heritability estimation reveals the genetic architecture of rare coding variants underlying complex traits
  Z. Liu, B. Fu, M. Jeong, P. Anand, **A. Anand**, S. Jang, A. Gorla, J. Zhu, P. Pajukanta, P. F. Palamara, N. Zaitlen, R. Border, S. Sankararaman · bioRxiv 2025 · under review at **Nature Genetics**

POSTERS

- Explaining potential epistasis in genomic data using symbolic representations of complex black box models
  A. Anand, P. Anand · Bruins in Genomics Summer Program, UCLA, August 2022

AWARDS, FELLOWSHIPS, AND GRANTS

Anthropic AI for Science Grant, Rare Disease Research, $50K in Claude credits — 2026
NIH T32 Predoctoral Fellowship, UCLA Genomic Analysis Training Program — 2026 – 2027
NHGRI T32 HG002536. One-year appointment, renewable.
Warren Alpert Computational Biology and AI Network Fellowship, UCLA — 2024
Attended the Computational Genomics Summer Institute in 2024 and 2025.
NSF REU Scholarship, Bruins in Genomics Summer Program, UCLA — 2022
Andy Grove Intel Scholarship, Intel — 2020
Dean's Honors List, UCLA — 2020 – 2024

TEACHING

Teaching Assistant, UCLA — 2025 – 2026

- Machine Learning Applications in Genetics (CS C124/224) — Spring 2026
- AI for Computational Genomics (CS CM121/221, ~120 students) — Spring 2025
  When the first half of the course was rebuilt around AI, co-wrote three new homework assignments and the discussion slides with the other TAs.
- Machine Learning Applications in Biomedicine (DS BMED 205) — Winter 2025

SOFTWARE

- **Inertia-1**: open-source wearable motion foundation models (co-developer).
- **FAME**: added the LD-block removal step and support for missing phenotypes.
- **pipeline-align-DNA**: added the HISAT2 aligner, plus fixes and container updates (8 merged pull requests).

INDUSTRY EXPERIENCE

Software Engineering Intern, LabCorp, Burlington, NC — Jun – Sep 2023

- Built reproducible Snakemake pipelines for high-volume genetic sequencing data.

PRESS

- "Researchers develop a new computational tool to understand how genetic interactions impact human traits" (on FAME). UCLA Health news release, December 2025.

SKILLS

Machine learning: PyTorch, self-supervised pretraining, time-series models, large-scale evaluation
Statistics and genetics: variance-component models, method of moments, heritability and genetic correlation, GWAS (plink2), LDSC, single-cell genetics
Engineering: Python, C++, SLURM/HPC, Snakemake, Nextflow, Docker, Git, Claude Code
Data: UK Biobank, NHANES, NHATS, OneK1K, GTEx