

# NIKHITA DAMARAJU

[nikhita@uw.edu](mailto:nikhita@uw.edu) • <https://nikhita-damaraju.github.io>

## EDUCATION

### UNIVERSITY OF WASHINGTON

Seattle, WA

*Doctor of Philosophy (PhD), Public Health Genetics*

September, 2022 - present

*Courses: Ethical, legal, social implications of Genomics; Principles of Health Metrics, Strategies of Health Promotion, Genetics and Law, Health Economics, Pharmacoeconomics, Pharmacogenomics, Qualitative Analysis*

### COLUMBIA UNIVERSITY MAILMAN SCHOOL OF PUBLIC HEALTH

New York, NY

*Master of Science (MS), Biostatistics – Statistical Genetics*

September, 2020 - May, 2022

*Relevant courses: Data Science, Machine Learning, Biostatistical Methods-I & II, Statistical Genetic Modelling, Population Genetics, Statistical Inference, Probability, Causal Inference, Probabilistic Graphical Models for Health Data, Epidemiology, Advance Statistical Genomics methods*

### INDIAN INSTITUTE OF TECHNOLOGY MADRAS

Chennai, TN, India

*Bachelor and Master of Science (B.S.-M.S.), Biological Sciences – Computational Biology*

August, 2015 – May, 2020

*Relevant Courses: Bioinformatics, Biostatistics, Computational Biology, Systems Biology, Quantitative and Population Genetics, Genomics and Proteomics, Immunology, Biochemistry, Microbiology, Molecular Biology, Cancer Biology, Biotechnology in Healthcare*

## SKILLS AND INTERESTS

### Computer Languages

R (rstats, shiny), Python (pandas, scikit-learn), SAS, C, C++, MATLAB

### Software

Bioinformatic analysis pipelines (short and long-read sequencing, RNA-seq (single-cell and bulk), ATAC-seq, CHIP-seq), ATLAS.ti

### Interests

Statistical Genetics, Biostatistics, Biomedical Data Science, Health Informatics

## HONORS, AWARDS AND GRANTS

- Building Bridges Grant with Dr. Danny Miller, Dr. Brian Shirts, UW Department of Lab Medicine & Pathology (2023)
- Scholarship recipient, 26<sup>th</sup> Summer Institute in Statistical Genetics, University of Washington (2021)
- Data Science Institute Scholarship, Columbia University (2020)
- Runners up and best software tool out of 15 teams, AI Genomics Hackathon, SVAI, San Francisco (2019)
- Khorana fellowship grant, Government of India, Indo-U.S. Science and Technology Forum (2018)
- Government of India DST INSPIRE fellow for five consecutive years given for best academic record, Indian Institute of Technology, Chennai (2015-2020)

## PROFESSIONAL EXPERIENCE

### UNIVERSITY OF WASHINGTON

Seattle, WA

*Graduate Research Assistant, Department of Laboratory Medicine*

January, 2023 - present

*Advisors: Dr. Brian Shirts, Dr. Danny Miller*

- Analyzing the performance of various variant calling algorithms on long read sequencing data.
- Determining the accuracy of phasing algorithms using data from different trios.

### UNIVERSITY OF WASHINGTON

Seattle, WA

*Graduate Research Assistant, School of Pharmacy*

March – September, 2023

*Advisor: Dr. Andy Stergachis*

- Performed statistical analysis for the cohort event monitoring of patients treated with a Tenofovir/Lamivudine/Dolutegravir (TLD) regimen in Mozambique.
- Characterized the types, rates, and risk factors for adverse events in the TLD Regimen using regression modelling.

## **ICAHN SCHOOL OF MOUNT SINAI**

*Department of Genetics and Genomic Sciences*

*Advisor: Dr. Paul O'Reilly*

New York, NY

February – August, 2022

- Developed a novel shrinkage methods for improving prediction of a polygenic risk score model on UK Biobank data
- Summarized key algorithmic considerations and biological insights in the form of a manuscript to be submitted

## **COLUMBIA UNIVERSITY, GENENTECH INC.**

*Data Science Institute Scholar 2020-2021, Herbert Irving Center For Cancer Research*

*Advisor: Dr. Alison Taylor*

New York, NY

October, 2020 – May, 2022

- Created a UNIX and R-based analysis pipeline on AWS consisting of copy number variant prediction algorithms on whole exome sequencing datasets to analyze differences between healthy and tumor samples.
- Quantified mutations (point and indels) stratified by chromosomal regions and genes across samples
- Worked on a collaborative study with Genentech Inc. to understand correlations between immunotherapy status and copy number variations in multiple tumor types

## **INFLAMMATIX INC.**

*Computational biology summer intern*

*Advisor: Dr. Yehudit Hasin, Dr. Yudong He*

Burlingame, CA

May – September, 2021

- Implemented a multi-cohort analysis framework on 20 publicly available datasets consisting of a heterogenous set of >2000 samples.
- Derived a gene expression signature to distinguish between healthy controls and Systemic Lupus Erythematosus (SLE), validated in an independent set with high accuracy.

## **INDIAN INSTITUTE OF TECHNOLOGY MADRAS**

*Dual Degree Thesis, Robert Bosch Center for Data Science and AI*

*Thesis advisor: Dr. Himanshu Sinha*

Chennai, TN, India

July 2019 – May 2020

- Assessed multiple gestational age (GA) estimation models on preterm birth prediction for a cohort of North-Indian women.
- Designed a regression-based model for GA estimation in first trimester with highest preterm labelling accuracy [1].
- Implemented machine learning approaches using ultrasound metrics to predict GA accurately in second and third trimesters. [2]
- Worked in close collaboration with a team of clinicians at the Translational Health Sciences Institute, Delhi as a part of Bill and Melinda Gates foundation grant, Grand Challenges of India grant and BIRAC grant.
- Presented research at multiple Data Science and AI academic conferences.

## **STANFORD UNIVERSITY**

*Summer Research Intern, Department of Biochemistry*

*Advisor: Dr. Julia Salzman*

Stanford, CA

May – August 2019

- Developed a generalized linear model using R to predict alternative splicing events consisting of 50 features using primate transcriptomic data combined with genomic information of key regulatory elements.
- Identified multiple SINE and LINE repeat aggregations conserved across six primate families.

## **STANFORD UNIVERSITY**

*Khorana summer fellowship intern, Department of Biochemistry*

*Advisor: Dr. Julia Salzman*

Stanford, CA

May – August 2018

- Designed a bioinformatic analysis pipeline for analyzing circRNAs in Epstein Barr virus infected human B-cell transcriptomic datasets downloaded from SRA to identify differences in lytic and latent infection stages of the virus.
- Selected as one of the 40 Khorana fellowship grant recipients among 1000 applicants to conduct research at the Salzman Lab, Department of Biochemistry.

## **TEACHING AND LEADERSHIP**

### **COLD SPRING HARBOR LABORATORY**

New York, NY

#### ***Teaching Assistant, Computational Genomics course***

November, 2023

- Helped in troubleshooting bioinformatic issues during the week long course focused on theory and practice of algorithms in computational biology.
- Covered fundamentals on “Haplotype Analysis” as one of course modules.
- Advised a group project on ABC analysis (RNA-seq, ATAC-seq and ChIP-Seq) for differential expression in maternal and fetal tissue data.

### **UNIVERSITY OF WASHINGTON**

Seattle, WA

#### ***Teaching Assistant, School of Public Health***

September, 2022 – June, 2023

- PHG303: Direct-to-Consumer Genetic Testing: Uses and Issues: Responsible for holding discussion sections based on case studies along with weekly office hours and grading.
- PHRMRA 545: Statistical Topics for Biomedical Regulatory Affairs Professionals: Responsible for grading homework, holding office hours, attending lecture and providing additional help with statistical methods.
- BIOST 511: Medical Biometry: Responsible for grading homework, holding discussion sections to enable students to practice course concepts and weekly office hours. I am also developed content for weekly discussion sections.
- PHG 200: Implications of Public Health Genomics for Society: Responsible for holding discussion sections comprising of debates based on articles related to course content along with weekly office hours.

### **COLUMBIA UNIVERSITY**

New York, NY

#### ***Teaching Assistant, Department of Biostatistics***

August – December, 2021

- P8105: Data Science: Responsible for grading homework assignments and mentoring student groups for the final project. I held weekly office hour sessions that involve debugging and explaining tidyverse/dplyr functions to students.
- P8104: Probability: Responsible for conceptualizing questions for weekly assignments, holding office hours to clarify theoretical concepts, grading homework and examinations.

### **INDIAN INSTITUTE OF TECHNOLOGY MADRAS**

Chennai, TN, India

#### ***Teaching Assistant***

July, 2019 – May, 2020

- Quantitative and Population Genetics: formulated questions for quizzes and end semester examinations along with grading of 2 online research seminars for over 20 students
- Synthetic Biology: responsible for assessing and grading research seminars for over 25 students

#### ***Founder and Head, Biotech Research Club***

September, 2017 – June, 2019

- Started a student organization dedicated at developing an interest in Synthetic biology among the students of IIT Madras where I was responsible for planning and conducting relevant talks by graduate students, experimental workshops to compliment theoretical understanding and journal clubs on dedicated topics.
- Curated over 10 articles for a science magazine called `Synkranti` as the chief editor

## **CONFERENCES AND POSTERS**

- 2023 Genome Informatics meeting, Cold Spring Harbor Laboratory, NY, short talk
- 2023 Stanford Genetics Conference on Structural Variants and DNA Repeats, short talk
- 2021 Initiative for Biological Systems Engineering 8th Workshop, talk
- 2020 7th RBCDSAI Workshop on Recent Progress in Data Science and AI, talk
- 2020 India EMBO symposium on Synthetic Biology, presentation and volunteer
- 2019 Initiative for Biological Systems Engineering 5th workshop, award for best poster and presentation
- 2019 Tata consultancy services research presentation competition, award for best poster and presentation
- 2016-2018 International Genetically Engineered Machine Competition, Boston
  - Gold medal, Foundational Advance track (2018)
  - Silver medal, Software track (2017) for published tool ChassiDex [5]
  - Silver medal, Measurement track (2016)

## **PEER-REVIEWED PUBLICATIONS**

- [1] Gadekar VP\*, **Damaraju N\***, et al. Development and external validation of Indian population-specific Garbhini-GA2 model for estimating gestational age in second and third trimesters. *The Lancet Regional Health - Southeast Asia*. 2024;0(0). doi:10.1016/j.lansea.2024.100362 (\*co- first authors)
- [2] **Damaraju N**, Miller AL, Miller DE. Long-Read DNA and RNA Sequencing to Streamline Clinical Genetic Testing and Reduce Barriers to Comprehensive Genetic Testing. *J Appl Lab Med*. 2024;9(1):138-150. doi:10.1093/jalm/jfad107
- [3] Vijayram R\*, **Damaraju N\***, Xavier A\*, et al. Comparison of first trimester dating methods for gestational age estimation and their implication on preterm birth classification in a North Indian cohort. *BMC Pregnancy and Childbirth*. 2021; 21(1):343. doi:10.1186/s12884-021-03807-4 (\*co- first authors)

## **PREPRINTS**

- [5] Kailash BP\*, Karthik D\*, Shinde M\*, **Damaraju N\***, et al. *ChassiDex: A Microbial Database Useful for Synthetic Biology Applications.*; 2019:703033. doi:10.1101/703033 (\*co- first authors)
- [6] **Damaraju N\***, Xavier A\*, et al. Development of Second and Third-Trimester Population-Specific Machine Learning Pregnancy Dating Model (Garbhini-GA2) Derived from the GARBH-Ini Cohort in North India.; 2021: 2021.10.02.21264450. doi:10.1101/2021.10.02.21264450 (\*co- first authors)

## **OTHER WRITING**

- [7] **Nikhita Damaraju**, “Bench-to-Bedside: A Dream or Reality?”, Apr 13, 2022, Columbia University Mailman School of Public Health Student Voices Blog
- [8] Ashley Xavier, Himanshu Sinha, **Nikhita Damaraju**, “Developing India-specific pregnancy dating model from Garbhini cohort”, AI for Social Good, Feb 9, 2021, Robert Bosch Centre for Data Science and AI (RBCDSAI)