

Douglas Yao

douglasyao@g.harvard.edu • (408) 510-4019 • douglasyao.github.io

EDUCATION

Harvard University

- Ph.D. in Systems, Synthetic, and Quantitative Biology

Jul 2018 – Oct 2023

University of California, Los Angeles

- B.S. in Molecular Biology (GPA: 3.82, magna cum laude) Sep 2014 – Jun 2018
- Additional coursework: Discrete Structures (Math 61, Grade: A), Algorithms and Complexity (CS 180, Grade: A), Probability (Stat 100A, Grade: A+), Mathematical Statistics (Stat 100B, Grade: A), Linear Algebra (Math 33A, Grade: A-), Computational Genetics (CS 124, Grade: A), Advanced Programming (PIC 10C, Grade: A), Machine Learning (CS 226, Grade: A-)

RESEARCH EXPERIENCE

Harvard University

Functional genomics technology development

Jan 2020 – Sep 2023

- Led a project to greatly increase the throughput of Perturb-seq, a type of high-throughput genetic assay that perturbs many different genes in cells using CRISPR and reads out gene expression profiles using single-cell RNA-sequencing. With a team of experimental biologists, reduced the cost of Perturb-seq by 90% over existing strategies using new experimental and computational techniques inspired by compressed sensing.
- Wrote first-author manuscript detailing this project, which is currently in press at Nature Biotechnology (impact factor: 68.1).

Statistical genetics research

Jul 2018 – Jan 2020

- Conceived and led a project developing a new statistical method to estimate heritability mediated by gene expression levels from human population-genetic data.
- Wrote first-author manuscript detailing this project, which is published in Nature Genetics (impact factor: 41.3). Paper cited >150 times since publication in 2020.

University of California, Los Angeles

Cancer genomics research

Feb 2016 – Jun 2018

- As an undergraduate, conceived and solely led a project applying linear mixed models to identify genes whose expression is associated with genomic instability in cancer
- Wrote first-author manuscript detailing this project, which is published in Scientific Reports.

SELECTED PUBLICATIONS

Yao DW, Binan L, Bezney J, Simonton B, Freedman J, Frangieh CJ, Dey K, Geiger-Schuller K, Eraslan B, Gusev A, Regev A, Cleary B. Scalable genetic screening for regulatory circuits using Compressed Perturb-seq. *Nature Biotechnology*, in press.

Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nature Genetics*. **52**, 626–633 (2020).

Yao DW, Balanis NG, Eskin E, Graeber TG. A linear mixed model approach to gene expression-tumor aneuploidy association studies. *Scientific Reports*. **9**, 1-8 (2019).

SKILLS

Programming languages: Python, R, Unix/Linux

Statistics/machine learning: Dimensionality reduction (PCA, NMF, UMAP), clustering, penalized regression (scikit-learn), generalized linear models (glmnet), convex optimization (cvxpy), matrix completion (softImpute), Bayesian statistical inference (Stan, Tensorflow Probability), deep learning (Keras, Tensorflow), variance component models