

# Neil Thomas

<https://people.eecs.berkeley.edu/~nthomas/>

1601 Allston Way  
Berkeley, CA 94703

nthomas@berkeley.edu  
(310) 699-7973

---

<b>EDUCATION</b>	<b>University of California, Berkeley</b> <b>PhD Student, Computer Science</b> <ul style="list-style-type: none"><li>My research focuses on developing machine learning methods for learning meaningful representations of proteins, with the aim of enabling applications in protein design, functional annotation, homology detection, and structure prediction.</li></ul>	2017 - Present
	University of California, Berkeley BS, Engineering Mathematics & Statistics, <i>High Honors</i> (GPA: 3.82)	2011 - 2015
	Independent University of Moscow, Russia Study-abroad program: <i>Math in Moscow</i>	Spring 2014
<b>SELECTED COURSEWORK</b>	Organic Chemistry, Microbial Ecology, Statistical Mechanics, Convex Optimization	
<b>PUBLICATIONS</b>	Nicholas Bhattacharya*, <b>Neil Thomas*</b> , Roshan Rao, Justas Dauparas, Peter K. Koo, David Baker, Yun S. Song, Sergey Ovchinnikov “Single Layers of Attention Suffice to Predict Protein Contacts” <i>bioRxiv</i> (2020)	
	Roshan Rao*, Nicholas Bhattacharya*, <b>Neil Thomas*</b> , Yan Duan, Xi Chen, John Canny, Pieter Abbeel, Yun S. Song “Evaluating Protein Transfer Learning with TAPE” <i>Advances in Neural Information Processing Systems 32 (NeurIPS 2019)</i> . Selected as <b>Spotlight Talk</b> . (2.4% of submissions)	
<b>WORK EXPERIENCE</b>	<i>Research Intern</i> Google Accelerated Sciences, Remote <ul style="list-style-type: none"><li>Working on improving machine learning methods for protein design.</li></ul>	Aug 2021 - Present
	<i>AI Resident</i> X (formerly Google X), Remote <ul style="list-style-type: none"><li>Part of an early-stage synthetic biology moonshot.</li></ul>	March 2021 - Aug 2021
	<i>Software Engineer</i> 23andMe, Mountain View, CA <ul style="list-style-type: none"><li>Implemented IBD (Identity by Descent) pipeline in Apache Spark to take advantage of data locality and parallelism. Relieved backlog of customer processing.</li><li>Automated imputation pipeline using Minimac3 and Luigi. Imputed over 1 million individuals on custom reference panel with improved accuracy.</li><li>Deployed to AWS to scale computation pipelines to handle massive customer demand.</li><li>Built haplotype phasing application using modified BEAGLE algorithm. Wrote Python wrappers for extracting and deploying legacy research C++ code.</li><li>Responsible for collaborative project planning - distilling input from multiple teams. Interviewed candidates. Mentored and onboarded new engineers.</li></ul>	Aug 2015 - June 2017
<b>TEACHING</b>	<i>Graduate Student Instructor</i> Mathematical Statistics (STAT 135), UC Berkeley <ul style="list-style-type: none"><li>Core upper division course for statistics majors. Topics include: parameter estimation, hypothesis testing, linear regression.</li><li>Led 3 weekly discussion sections, totalling 40 students.</li></ul>	Fall 2020
<b>AWARDS</b>	<i>NIH Genomics Training Grant</i>	2018 - 2020
	<i>Best Lightning Talk</i> - UC Berkeley Computational Biology Retreat	2019
	<i>Winner</i> - 23andMe Hackathon	2016

- Part of a 5-person team that, in 1.5 days, built an interactive way to explore 23andMe genetic reports using the human body.

*Honorable Mention* (top 40%) - COMAP Mathematical Contest in Modeling 2015

- **Neil Thomas**, Lukas Whaley-Mayda, Miles Rusch. “Leveraging the Criticality of Outbreaks to Eradicate Ebola.” February 2015. Simulated vaccine effect on an Ebola outbreak using site-percolation model.

## RESEARCH EXPERIENCE

*SURF/Rose Hills Undergraduate Research Fellow* 2014

Dept. of Statistics, University of California, Berkeley

- Designed and implemented computational experiments to test the performance of a demographic inference algorithm on simulated genetic data with phasing errors.
- **Neil Thomas**, Geno Guerra, Yun S. Song. “Assessing the Effect of Haplotype Phasing on Demographic Inference.” Student poster presented at SURF research conference. August 2014. Awarded *First Prize*.

*SURF/Rose Hills Undergraduate Research Fellow* 2013

Dept. of Integrative Biology, University of California, Berkeley

- Designed and implemented an experimental setup for testing marine larval response to linear fluid accelerations. Analyzed data linking accelerations to behavioral response.
- **Neil Thomas**, Rachel Pepper, Dorian Liepmann, M.A.R. Koehl. “Simple microfluidic-inspired devices for observation of small aquatic organisms in controllable fluid flow: design and implementation.” Student poster presented at the 66th Annual Meeting of the APS Division of Fluid Dynamics. November 2013.

## LEADERSHIP & SERVICE

*Board Representative*, Berkeley Student Cooperative Aug 2020 - March 2021

Member of executive board of \$14M/year nonprofit that provides affordable housing to students. Crafted policy and planned budget as part of Capital and Finance Committee.

*Network Manager*, The Convent, Berkeley Student Cooperative Jan 2018 - Present

Manage IT infrastructure for 25-person dwelling.

*Organizer*, Protein ML Reading Group Nov 2019 - Dec 2020

Weekly online reading group focused on machine learning for biology. Facilitated discussion and recruited speakers.

## LANGUAGES

Python, R, Java

## TOOLS

AWS, Ansible, Apache Spark, Azure, Git, Jenkins, JIRA, L<sup>A</sup>T<sub>E</sub>X, MySQL, Packer, PyTorch, Tensorflow, Vagrant

## FOREIGN LANGUAGES

Russian, *Fluent*

## HOBBIES

Ultimate Frisbee, Piano, Cycling, Hiking, Rap, Improv, Dunking<sup>†</sup>

<sup>†</sup> - *in progress*