

Gregory Darnell

Research Scientist
Health AI Division of AI/ML
Apple Inc.

Phone: (617) 851 9752
Email: gdarnell@apple.com
Home: <http://www.gregdarnell.com/>

Education

2014–2019	Ph.D. Quantitative and Computational Biology	Princeton University
2015	M.A. Quantitative and Computational Biology	Princeton University
2013–2014	Ph.D. Student Computational Biology & Bioinformatics	Duke University
2008–2013	B.S. Computer Science, <i>minor</i> Bioinformatics	University of California, Los Angeles

Postdoctoral Fellowship Research

My postdoctoral research builds upon my PhD goal of creating machine learning models that yield insight into vast genomic data. I am co-advising on projects at the intersection of machine learning, statistics, and genomics. These projects include: [1] biologically annotated neural networks with sparse priors for robust variable selection; [2] a framework that connects probabilistic autoencoders with deterministic autoencoders to mitigate mode collapse; [3] a new estimator for pathway-based gene scores to determine the aggregated influence of pathway variants on complex traits. Additionally, I am leading a project to perform imputation of non-linear interactions between genetic variants using only linear projections. I prove in the setting of additively-coded genetic mutations, higher moments of the variables allow this imputation to be possible.

Publications

Dissertations

- [1] **Gregory Darnell**. Associations and Confounding in High-Dimensional Genomics, Princeton University, Ann Arbor, 2019. ProQuest, <https://search.proquest.com/pqdtglobal/docview/2245698877/D7786861C23647ABPQ/1>

My dissertation research develops novel machine learning and statistical techniques for estimating genotype-phenotype associations across high-dimensional and multi-modal traits such as: autoimmune diseases, gene expression, and tissue morphology. Robust estimation of effect sizes and variable selection in this context is crucial – therapeutic techniques for disease are predicated on an understanding of the true genetic architecture of complex traits. My dissertation results have shown that by properly controlling for confounding factors, it is possible to maintain interpretability, increase statistical power, and gain insight into the genetic architecture of complex traits.

Refereed Journals

*Indicates equal authorship.

- [1] Li-Fang Cheng, Bianca Dumitrascu, **Gregory Darnell**, Corey Chivers, Michael Draugelis, Kai Li, Barbara E Engelhardt. “Sparse multi-output Gaussian processes for online medical time series prediction” *BMC Medical Informatics and Decision Making*. July, 2020
Efficient prediction of sparsely sampled biomarkers in a hospital setting using Gaussian Process regression.
- [2] Jordan Ash*, **Gregory Darnell***, Daniel Munro, Barbara E Engelhardt. “Joint analysis of gene expression levels and histological images identifies genes associated with tissue morphology.” *accepted. Nature Communications*. May, 2019
First computational pipeline to estimate genotype-image associations using neural networks.
- [3] Bianca Dumitrascu, **Gregory Darnell**, Julien Aroyles, Barbara E Engelhardt. “Statistical tests for detecting variance effects in quantitative trait studies.” *Bioinformatics*. July, 2018
A novel regression framework to estimate genetic variant effect on the variance of a phenotype.
- [4] **Gregory Darnell**, Stoyan Georgiev, Sayan Mukherjee, Barbara E Engelhardt. “Adaptive Randomized Dimension Reduction on Massive Data.” *Journal of Machine Learning Research (JMLR)*. 18(140):1-30, 2017
A solution to large inverse problems in linear mixed models using random projections.

- [5] **Gregory Darnell**, Dat Duong, Buhm Han, Eleazar Eskin. "Incorporating Prior Information into Association Studies." *Bioinformatics*. 28(12):i47-53, *Special Issue of the Nineteenth International Conference on Intelligent Systems in Molecular Biology (ISMB-2012)* Long Beach, CA: July 15-27, 2012
Increasing power in association studies by re-prioritizing variants based on prior information.

Archival

- [1] Pinar Demetci, Wei Cheng, **Gregory Darnell**, Xiang Zhou, Sohini Ramachandran, Lorin Crawford. "Multi-scale genomic inference using Biologically Annotated Neural Networks." *bioRxiv preprint* <https://www.biorxiv.org/content/10.1101/2020.07.02.184465v2>. July, 2020
- [2] Wei Cheng, **Gregory Darnell**, Sohini Ramachandran, Lorin Crawford. "Generalizing Variational Autoencoders with Hierarchical Empirical Bayes." *arXiv preprint* <https://arxiv.org/abs/2007.10389>. July, 2020

Professional Experience

- | | |
|-----------|--|
| 2020- | Research Scientist in Health AI Division of AI/ML, Apple Inc. (June -)
Applied Machine Learning Research for Health |
| 2019–2020 | Institute Postdoctoral Fellow, ICERM, Brown University (Sept - June)
Institute for Computational and Experimental Research in Mathematics
Mentors: Professors Sohini Ramachandran, Lorin Crawford, Sigal Gottlieb, Yanlai Chen
Program: Model and dimension reduction in uncertain and dynamic systems |
| 2017 | Research Internship, Microsoft Research New England (Sept - Dec)
Mentors: Dr. Nicolo Fusi
Project: Automated machine learning using parameter-free optimization |
| 2016 | Research Internship, Microsoft Research New England (June - Sept)
Mentors: Professor Jennifer Listgarten and Dr. Nicolo Fusi
Project: Output warping for improved prediction in machine learning models |
| 2012 | Research Assistant, Rockefeller University (June - Sept)
Mentors: Professors Robert Darnell and Chaolin Zhang
Projects: Accurate and efficient RNA-Seq read mapping &
Prediction of binding specificity of RNA-binding proteins |
| 2009 | Research Assistant, Harvard School of Public Health (June - Sept)
Mentor: Professor Alkes Price
Project: Geographical ancestry prediction using principal components analysis on genotypes |
| 2008–2013 | Research Assistant, University of California, Los Angeles
Mentor: Professor Eleazar Eskin
Projects: Improving power in association study methods in Bioinformatics &
Data compression schemes in computational genetics |

Invited Talks

- “Finding missing epistasis: Partitioning marginal epistasis deconvolves nonlinear interactions from additive effects in GWA summary statistics.” American Society of Human Genetics (ASHG) 2020, Platform Session. San Diego, California, October 27, 2020. (now virtual).
- “The path to exciting research directions and fulfilling mentorship in Bioinformatics” UCLA Bioinformatics Graduation Keynote Speaker. Los Angeles, California. June 7, 2017
- “Winner’s Curse in Quantitative Genomics Studies” *Biological Data Science*. Cold Spring Harbor, NY. Oct 28, 2016
- “Winner’s Curse in Quantitative Genomics Studies” *New York Area Population Genomics Workshop 2016*. Princeton University, NJ. Jan 21, 2016

Teaching

Summer 2016	Computer Science: Undergraduate Mentorship Project: Analyzing the link between ancestry and morphology in sea urchin embryo images	Princeton University
Fall 2015	COS 513: Foundations of Probabilistic Modeling Assistant in Instruction	Princeton University
Fall 2014	COS 375: Computer Architecture and Organization Assistant in Instruction	Princeton University
Spring 2012	Computer Science CM124. Computational Genetics Grader, Homework office hours, Test administration	University of California, Los Angeles

Technical Reviewing and Committees

2020 Program Committee Member NeurIPS (Conference on Neural Information Processing Systems).

2019, 2020 Program Committee Member IEEE International Conference on Bioinformatics and Biomedicine.

2016, 2017, 2018 Reviewer for the journal of *Bioinformatics*, Oxford Academics.

Academic Awards & Accreditations

- 2020 ASHG/Charles J. Epstein Award for Excellence in Human Genetics Research Postdoctoral SemiFinalist
- Selected for Institute Postdoctoral Fellowship at ICERM, 2019 (Institute for Computational and Experimental Research), Brown University
- Selected for the 2017 NBA Hackathon, New York, NY
- Travel Fellowship to ISMB 2012, Long Beach, CA
- COMPTIA A+, Network+, Microsoft Certified Professional, 2006

Leadership

- Student leader of statistics/machine learning reading group at Princeton.
- Co-founder and president of the UCLA Powerlifting Club Sport Team.

Skills

Python, PyTorch, R, Linux/Unix