

# Alejandro Ochoa

Duke Center for Statistical Genetics and Genomics  
Department of Biostatistics and Bioinformatics  
Fitzpatrick Center for Interdisciplinary Engineering, Medicine and Applied Sciences (CIEMAS)  
101 Science Dr, Room 2177C  
Durham, NC 27705  
[alejandro.ochoa@duke.edu](mailto:alejandro.ochoa@duke.edu)  
<https://ochoalab.github.io/>

## Education

- 2013 PhD in Molecular Biology, Princeton University  
Dissertation: *Protein domain prediction using context statistics, the false discovery rate, and comparative genomics, with application to Plasmodium falciparum*  
Advisers: Mona Singh (Computer Science) and Manuel Llinás (Molecular Biology)
- 2006 BS in Biology and Mathematics, Massachusetts Institute of Technology

## Awards & Honors

- 2020 Whitehead Scholar, Whitehead Charitable Foundation. For Duke junior faculty with exceptional potential for research and research training in the biomedical sciences.
- 2008 NSF Graduate Research Fellowship. Computational Biology.
- 2008 Ford Foundation Diversity Fellowship Predoc. Comp. (Declined, listed as Honorable Mention.)
- 2006 MIT Department of Biology Merck Prize. Awarded to a graduating senior for outstanding research and academic performance in biophysics or bioinformatics.
- 2001 Mexican Math Olympiad Gold Medal. Awarded to top 15 competitors nationally.

## Employment

- 2018-now *Assistant Professor*, Duke Center for Statistical Genetics and Genomics, Department of Biostatistics and Bioinformatics, Duke University
- 2013-2018 *Postdoctoral Research Associate* at John D Storey's group, Lewis-Sigler Institute for Integrative Genomics, and Center for Statistics and Machine Learning, Princeton University

## Publications

### JOURNAL ARTICLES

- 2023 Rachel Kate Cason, Eileen T Chambers, Tiffany Tu, Megan Chryst-Stangl, Kinsie Huggins, Brandon M Lane, **Alejandro Ochoa**, Annette M Jackson, Rasheed Gbadegesin. Genetic risk variants for childhood nephrotic syndrome and corticosteroid response. *Frontiers in Pediatrics*. 11:1248733. PMID [37868272](https://pubmed.ncbi.nlm.nih.gov/37868272/).

- 2023 Yiqi Yao, **Alejandro Ochoa**. Limitations of principal components in quantitative genetic association models for human studies. *eLife*. 79238. PMID [37140344](#).
- 2023 Zhuoran Hou, **Alejandro Ochoa**. Genetic association models are robust to common population kinship estimation biases. *Genetics*. iyado30. PMID [36843304](#).
- 2021 Brandon M Lane, Megan Chryst-Stangl, Guanghong Wu, Mohamed Shalaby, Sherif El Desoky, Claire Middleton, Kinsie Huggins, Amika Sood, **Alejandro Ochoa**, Andrew F Malone, Ricardo Vancini, Sara Miller, Gentzon Hall, So Young Kim, David Howell, Jameela A Kari, Rasheed Gbadegehin. CLVS1, a candidate gene for steroid sensitive nephrotic syndrome, regulates podocyte oxidative stress and endocytosis. *JCI Insight*. 7(2) e152102. PMID [34874915](#).
- 2021 Brian I Shaw, **Alejandro Ochoa**, Cliburn Chan, Chloe Nobuhara, Rasheed Gbadegehin, Annette M Jackson, Eileen T Chambers. HLA Loci and Recurrence of Focal Segmental Glomerulosclerosis In Pediatric Kidney Transplantation. *Transplant Direct*. 7(10) e748. PMID [34476293](#).
- 2021 Young-Sook Kim, Graham D Johnson, Jungkyun Seo, Alejandro Barrera, Thomas N. Cowart, William H Majoros, **Alejandro Ochoa**, Andrew S Allen, Timothy E Reddy. Correcting signal biases and detecting regulatory elements in STARR-seq data. *Genome Res*. 31(5) 877-889. PMID [33722938](#).
- 2021 **Alejandro Ochoa**, John D Storey. Estimating  $F_{ST}$  and kinship for arbitrary population structures. *PLoS Genet*. 17(1) e1009241. PMID [33465078](#).
- 2017 **Alejandro Ochoa**, Mona Singh. Domain prediction with probabilistic directional context. *Bioinf*. 33(16) 2471-8. PMID [28407137](#).
- 2016 Simon A Cobbold, Joana M Santos, **Alejandro Ochoa**, David H Perlman, Manuel Llinás. Proteome-wide analysis reveals widespread lysine acetylation of major protein complexes in the malaria parasite. *Sci Rep*. 2016;6:19722. PMID [26813983](#).
- 2015 **Alejandro Ochoa**, John D Storey, Manuel Llinás, Mona Singh. Beyond the  $E$ -value: stratified statistics for protein domain prediction. *PLoS Comput Biol*. 11 e1004509. PMID [26575353](#).
- 2013 Moriah L Szpara, Derek Gatherer, **Alejandro Ochoa**, Benjamin Greenbaum, Aidan Dolan, Rory J Bowden, Lynn W Enquist, Matthieu Legendre, Andrew J Davison. Evolution and diversity in human herpes simplex virus genomes. *J Virol*. 88:1209-27. PMID [24227835](#).
- 2011 **Alejandro Ochoa**, Manuel Llinás, Mona Singh. Using context to improve protein domain identification. *BMC Bioinformatics*. 12:90. PMID [21453511](#).
- 2007 Gevorg Grigoryan, **Alejandro Ochoa**, Amy E Keating. Computing van der Waals energies in the context of the rotamer approximation. *Proteins*. 68(4) 863-78. PMID [17554777](#).

#### MANUSCRIPTS IN SUBMISSION

- 2019 **Alejandro Ochoa**, John D Storey. New kinship and  $F_{ST}$  estimates reveal higher levels of differentiation in the global human population. Preprint: <https://doi.org/10.1101/653279>
- 2016 **Alejandro Ochoa**, John D Storey.  $F_{ST}$  and kinship for arbitrary population structures I: Generalized definitions. Preprint: <https://doi.org/10.1101/083915>

#### ACKNOWLEDGMENTS

- 2019 Irineo Cabreros, John D Storey. A Likelihood-Free Estimator of Population Structure Bridging Admixture Models and Principal Components Analysis. *Genetics*. 212(4) 1009-29. PMID [31028112](#).

- 2016 Prem Gopalan, Wei Hao, David M Blei, John D Storey. Scaling probabilistic models of genetic variation to millions of humans. *Nat Genet.* 48(12) 1587-90. PMID [27819665](#).
- 2014 Anton V Persikov, Mona Singh. De Novo Prediction of DNA-Binding Specificities for Cys2His2 Zinc Finger Proteins. *Nucleic Acids Res.* 42(1) 97-108. PMID [24097433](#).

## Distributed software

### MAJOR PACKAGES

- 2021-2023 simfam: Simulate and Model Family Pedigrees With Structured Founders. R, C++. Available on [CRAN](#) and <https://github.com/OchoaLab/simfam>.
- 2019-2023 genio: Genetics Input/Output Functions. R, C++. Available on [CRAN](#) and <https://github.com/OchoaLab/genio>.
- 2019-2023 simtrait: Simulate Complex Traits from Genotypes. R. Available on [CRAN](#) and <https://github.com/OchoaLab/simtrait>.
- 2017-2023 popkin: Estimate Kinship and FST under Arbitrary Population Structure. R, C++. Available on [CRAN](#) and <https://github.com/StoreyLab/popkin>.
- 2017-2023 bnpsd: Model and Simulate Admixed Populations. R. Available on [CRAN](#) and <https://github.com/StoreyLab/bnpsd>.
- 2014-2020 dPUC2: Domain Prediction Using Context, Version 2. Perl, C. <https://github.com/alexviia/dpuc2>.
- 2014-2020 DomStratStats: Domain Stratified Statistics (q-values and local FDRs). Perl. <https://github.com/alexviia/DomStratStats>.
- 2014-2019 RandProt: High-order Markov random models for protein sequences. Perl. <https://github.com/alexviia/RandProt>.

### MINOR PACKAGES

- 2021-2023 simgenphen: Simulate Genotypes and Phenotypes. R. <https://github.com/OchoaLab/simgenphen>.
- 2021-2023 genbin: R wrappers for binaries in genetics. R. <https://github.com/OchoaLab/genbin>.
- 2020-2023 ligera: Light GENetic Robust Association. R, C++. <https://github.com/OchoaLab/ligera>.
- 2019-2022 popkinsuppl: Supplement to "popkin" package. R. <https://github.com/OchoaLab/popkinsuppl>.

### PAPER REPOSITORIES

- 2021-2023 bias-assoc-paper: Kinship bias association project. R, bash, LaTeX, markdown. <https://github.com/OchoaLab/bias-assoc-paper>.
- 2019-2023 pca-assoc-paper: PCA association project. R, bash, LaTeX, markdown. <https://github.com/OchoaLab/pca-assoc-paper>.
- 2019-2023 data: Instructions for real data processing shared across projects. R, bash, Perl, markdown. <https://github.com/OchoaLab/data>.

2019 human-differentiation-manuscript: Human differentiation analysis. R, bash, markdown.  
<https://github.com/StoreyLab/human-differentiation-manuscript>.

## Invited Talks

2024 *The 12th RECOMB Satellite Workshop on Computational Methods in Genetics*. Stata Center, Massachusetts Institute of Technology, Cambridge, MA. 2024-04-28. Presented by Tiffany Tu.

2024 *Princeton QCB seminar*. Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ. 2024-04-22.

2023 *Duke Computational Biology and Bioinformatics (CBB) retreat*. Holston Presbytery Camp and Retreat Center, Banner Elk, NC. 2023-11-18.

2023 *NIH Centers for Excellence in Genomics Science (CEGS) Annual Meeting*. Kimmel Center, New York University, New York, NY. 2023-10-11.

2023 *HLA and KIR region genomics in immune-mediated diseases consortium (virtual)*. National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD. 2023-06-16.

2023 *HCHS/SOL (Hispanic Community Health Study / Study of Latinos) Genetics Special Interest Group (Virtual)*. Department of Biostatistics, Harvard University, Cambridge, MA. 2023-06-21.

2023 *NIH Centers for Excellence in Genomics Science (CEGS) Virtual Outreach Meeting*. The Jackson Laboratory, Bar Harbor, ME. 2023-04-20.

2023 *American Mathematical Society (AMS) 2023 Spring Southeastern Sectional Meeting*. Georgia Institute of Technology, Atlanta, GA. 2023-03-19.

2022 *BERD Core Seminar*. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2022-08-26.

2022 *Latinx in the Mathematical Sciences Conference 2022 (LatMath2022)*. Institute for Pure and Applied Mathematics, University of California, Los Angeles, Los Angeles, CA. 2022-07-08.

2022 *Duke University Program in Genetics and Genomics (UPGG) retreat*. Suite 4, Durham, NC. 2022-06-11.

2022 *HLA and KIR region genomics in immune-mediated diseases consortium (virtual)*. National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD. 2022-04-28.

2021 *Orientation for MB and PhD students*. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2021-08-19.

2021 *HLA and KIR region genomics in immune-mediated diseases consortium (virtual)*. National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD. 2021-04-06.

2021 *Statistics seminar*. Department of Mathematics and Statistics, Washington State University, Pullman, WA. 2021-03-10.

2021 *Master of Biostatistics Virtual Visit Day*. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2021-02-19.

2020 *American Society of Human Genetics (ASHG) Annual Meeting*. Virtual. 2020-10-27. Presented by Amika Sood.

2020 *University Program in Genetics and Genomics (UPGG) seminar*. Room 147, Nanaline Duke building, Duke University, Durham, NC. 2020-02-18.

- 2019 *Duke Computational Biology and Bioinformatics (CBB) retreat.* Holiday Inn Resort, Wrightsville Beach, NC. 2019-09-28.
- 2019 *Computational Biology and Bioinformatics (CBB) seminar.* French Family Science Center, Duke University, Durham, NC. 2019-02-04.
- 2019 *4th Mexico Population Genomics Meeting.* Amoxcalli Complex, Faculty of Sciences, University City, National Autonomous University of Mexico, Mexico City, DF, Mexico. 2019-01-11.
- 2018 Population Biology seminar, Department of Biology, Duke University, Durham, NC. 2018-09-27.
- 2018 Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2018-02-21.
- 2018 Department of Biostatistics, Johns Hopkins University, Baltimore, MD. 2018-01-22.
- 2017 Department of Biology, University of Richmond, Richmond, VA. 2017-11-10.
- 2017 *Princeton Research Day.* Princeton University, Princeton, NJ. 2017-05-11.
- 2017 Department of Genetics, University of North Carolina, Chapel Hill, NC. 2017-02-13.
- 2016 *New York Area Population Genomics Workshop 2016.* Princeton University, Princeton, NJ. 2016-01-21.
- 2015 *Probabilistic Modeling in Genomics (ProbGen) conference.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 2015-10-15.
- 2013 *Telepresentation for Yun Song's group.* UC Berkeley, Berkeley, CA. 2013-04-17.
- 2013 *Biological sequence analysis and probabilistic models (ProbGen) conference.* HHMI Janelia Farm, Ashburn, VA. 2013-03-25.
- 2013 National Center for Biotechnology Information (NCBI), National Institutes of Health (NIH), Bethesda, MD. 2013-02-25.
- 2012 *Recruiting conference.* Department of Computer Science, Princeton University, Princeton, NJ. 2012-03-01.
- 2004 Rotary Club Paso del Norte, Ciudad Juarez, CHIH, Mexico. 2004-08-26.

## Conference posters

- 2024 Tiffany Tu, **Alejandro Ochoa**. The presence of cryptic relatedness creates inflation in meta-analyses of genome-wide association studies. *The 12th RECOMB Satellite Workshop on Computational Methods in Genetics, and 28th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*. Stata Center, Massachusetts Institute of Technology, Cambridge, MA. 2024-04-28. Presented by Tiffany Tu.
- 2023 Tiffany Tu, Rasheed Gbadegesin, Adebawale Adeyemo, **Alejandro Ochoa**. A test of genotyping platform bias for multiethnic case/control association studies merging external controls. *American Society of Human Genetics (ASHG) Annual Meeting*. Walter E. Washington Convention Center, DC. 2023-11-04. Presented by Tiffany Tu.
- 2023 Ratchanon Pornmongkolsuk, **Alejandro Ochoa**. Single-locus imputation of ancient African DNA using novel regression-based approach. *American Society of Human Genetics (ASHG) Annual Meeting*. Walter E. Washington Convention Center, DC. 2023-11-03. Presented by Ratchanon Pornmongkolsuk.
- 2023 Shannon Clarke, Makenzie Beaman, Yuncheng Duan, Apoorva Iyengar, Revathy Venukuttan, Eden Harris, Allison Ashley-Koch, Rasheed Gbadegesin, Opeyemi Olabisi, William H Majoros, Timothy E Reddy, **Alejandro Ochoa**. Duke Genomic Scholars Program: Providing Accessible Genomic Training for a Diverse Workforce. *NIH Centers for Excellence in Genomics Science (CEGS)*

*Annual Meeting*. Kimmel Center, New York University, New York, NY. 2023-10-11. Presented by Shannon Clarke.

Zhuoran Hou, **Alejandro Ochoa**. Genetic association models are robust to common population kinship estimation biases. *American Society of Human Genetics (ASHG) Annual Meeting*. Convention Center, Downtown, Los Angeles, CA. 2022-10-27. Presented by Zhuoran Hou.

**Alejandro Ochoa**, Yuncheng Duan, Revathy Venukuttan, Shannon Clarke, Timothy E Reddy. Development of Genomic Resource Modules for a Diverse Workforce. *NIH Centers for Excellence in Genomics Science (CEGS) Annual Meeting*. The Mary Duke Biddle Trent Semans Center for Health Education, Duke University, Durham, NC. 2022-10-19.

**Alejandro Ochoa**, Amika Sood. Joint inference of admixture and population history from the genetic covariance structure. *Probabilistic Modeling in Genomics (ProbGen) conference*. Virtual. 2021-04-14.

**Alejandro Ochoa**, John D Storey. New kinship and  $F_{ST}$  estimates applied to the global human population. *American Society of Human Genetics (ASHG) Annual Meeting*. Virtual. 2020-10-27.

**Alejandro Ochoa**, John D Storey. Relatedness and Differentiation in Arbitrary Population Structures. *4th Mexico Population Genomics Meeting*. Amoxcalli Complex, Faculty of Sciences, University City, National Autonomous University of Mexico, Mexico City, DF, Mexico. 2019-01-11.

**Alejandro Ochoa**, John D Storey. Relatedness and Differentiation in Arbitrary Population Structures. *Probabilistic Modeling in Genomics (ProbGen) conference*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 2018-11-05.

**Alejandro Ochoa**, John D Storey. Relatedness and Differentiation in Arbitrary Population Structures. *Population, Evolutionary and Quantitative Genetics Conference*. Madison Concourse Hotel, Madison, WI. 2018-05-15.

**Alejandro Ochoa**, John D Storey, Srikanth Gottipati, Shashank Rohatagi, Andrew Forbes, Deborah Profit, Raymond Sanchez, Margaretta Nyilas, William Carson. Mixed Effects Modeling of Placebo Response Across 8 Placebo-controlled Aripiprazole Trials Spanning 20 Years in Acutely Relapsed Schizophrenia Patients. *Neuroscience Education Institute Psychopharmacology Congress*. Broadmoor Convention Center, Colorado Springs, CO. 2016-11-03.

Srikanth Gottipati, **Alejandro Ochoa**, Shashank Rohatagi, Andrew Forbes, Deborah Profit, Raymond Sanchez, Margaretta Nyilas, William Carson. Canonical Loadings of PANSS Subscales Show Differential Placebo and Aripiprazole Drug Responses in Schizophrenia Patients. *Neuroscience Education Institute Psychopharmacology Congress*. Broadmoor Convention Center, Colorado Springs, CO. 2016-11-03. Presented by Srikanth Gottipati.

**Alejandro Ochoa**, John D Storey.  $F_{ST}$  generalized for arbitrary population structures. *John W. Tukey 100th Birthday Celebration conference*. Center for Statistics and Machine Learning, Princeton University, Princeton, NJ. 2015-09-18.

**Alejandro Ochoa**, John D Storey, Manuel Llinás, Mona Singh. Forget the  $E$ -value: family-based  $q$ -values for protein domain prediction, and empirical error detection. *Biological sequence analysis and probabilistic models (ProbGen) conference*. HHMI Janelia Farm, Ashburn, VA. 2013-03-25.

**Alejandro Ochoa**, Manuel Llinás, Mona Singh. Using context to predict protein domains across diverse organisms. *Research in Computational Molecular Biology (RECOMB) 3rd Joint Conference on Systems Biology, Regulatory Genomics, and DREAM Challenges*. Columbia University, New York, NY. 2010-11-09.

# Teaching

## DUKE UNIVERSITY

- 2022-2023 Human Genetics. (Co-Lecturer Spring 2022-2023)
- 2021-2023 UPGEN 778A-F Mini course: Genetic population structure and relatedness (Fall 2021-2023)
- 2019-2023 SIBS lecture on genetic association studies (Summer 2019, 2021-2023)
- 2019 BIOS 900 Current Problems in Biostatistics. Special lecture on kinship and  $F_{ST}$
- 2019 BIOS 710 Statistical Genetics and Genetic Epidemiology. Special lecture on kinship and  $F_{ST}$

## PRINCETON UNIVERSITY

- 2017 Intro to Genomics and Comp Bio. Quant Comp Bio, Comp Sci. (Co-Lecturer Fall 2017)
- 2016-2017 Discussion Leader. Summer Undergraduate Research Program (Summer 2016, 2017)
- 2014-2017 Statistical Treatment of Data. Mol Bio, Quant Comp Bio. (Workshop Fall 2014-2017)
- 2011 Intro to Genomics and Comp Mol Bio. Mol Bio, Comp Sci. (TA Fall 2011)
- 2008 Core Laboratory. Mol Bio. (TA Spring 2008)

## MASSACHUSETTS INSTITUTE OF TECHNOLOGY

- 2005 Calculus. OME Project Interphase (TA Summer 2005)
- 2003-2004 Multivar Calculus and Diff Eqs. OME Seminar XL (Tutor Fall 2003, Spring 2004)
- 2003-2004 Calculus. MITE2S Program (TA Summer 2003, 2004)

# Mentoring

## CURRENT TRAINEES - PRIMARY ADVISER

- 2020-now Tiffany Tu. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. Computational Biology and Bioinformatics Program, Duke University.
- 2022-now Ratchanon "RP" Pornmongkolsuk. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. University Program in Genetics and Genomics, Duke University.
- 2022-now Zhuoran Hou. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- 2023-now Danielle Mensah. Undergraduate. Research Independent Study. Ochoa Laboratory. Department of Computer Science, Duke University.

## CURRENT TRAINEES - SECONDARY ADVISER

- 2019-now Xue "Scarlett" Zou. PhD. Dissertation committee. Allen Laboratory. Computational Biology and Bioinformatics Program, Duke University.
- 2021-now Cymfenee Dean-Phifer. PhD. Graduate rotation, Dissertation committee. Goldberg Laboratory. Computational Biology and Bioinformatics Program, Duke University.
- 2022-now Changxin Wan. PhD. Dissertation committee. Ji Laboratory. Computational Biology and Bioinformatics Program, Duke University.
- 2023-now Bide "Peter" Chen. PhD. Dissertation committee. Goldberg Laboratory. University Program in

Genetics and Genomics, Duke University.

- 2023-now Jinting Justin Liu. Master's. Master's project committee. Majoros Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- 2023-now Gabriel Kennedy. PhD. Dissertation committee. Goldberg Laboratory. University Program in Genetics and Genomics, Duke University.
- 2023-now Anvita Kulshrestha. PhD. Dissertation committee. Ashley-Koch Laboratory. University Program in Genetics and Genomics, Duke University.
- 2024-now Jennifer Drucker Varner. Fellow. Scholarship Oversight Committee. Gbadegesin Laboratory. Department of Pediatrics, Duke University.

#### PAST TRAINEES - PRIMARY ADVISER

- 2019-2020 Yiqi Yao. Master's. BCTIP internship, Master's project adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now Senior Business Analyst at BenHealth.
- 2019-2022 Amika Sood. Postdoctoral. Postdoctoral adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now Staff Scientist at Complex Carbohydrate Research Center (CCRC), The University of Georgia.
- 2020-2021 Zhuoran Hou. Master's. BCTIP internship. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now PhD Student at Duke University, BB.
- 2020-2022 Jiajie Shen. Master's. Research, Master's project adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- 2021-2022 Emmanuel Mokel. Undergraduate. Research Independent Study. Ochoa Laboratory. Department of Statistical Science, Duke University.

#### PAST TRAINEES - SECONDARY ADVISER

- 2009 Neo Christopher Chung. PhD. Graduate rotation. Llinás Laboratory. Quantitative Computational Biology Program, Princeton University. Now Adjunct Faculty at University of Warsaw.
- 2010 Jeremy Bigness. PhD. Graduate rotation. Singh Laboratory. Quantitative Computational Biology Program, Princeton University. Now PhD Student at Brown University.
- 2011 Sebastian Nasamu. Undergraduate. Summer project. Llinás Laboratory. Department of Molecular Biology, Princeton University. Now Post-doctoral Research Fellow at Johns Hopkins Bloomberg School of Public Health.
- 2019 Yuncheng Duan. PhD. Graduate rotation. Ochoa Laboratory. Department of Biology, Duke University. Now PhD Student at Duke University, Allen Lab.
- 2019-2020 Shengyu Li. Master's. Master's project committee. Allen Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now PhD Student at Duke University, CBB.
- 2020-2022 Iman Hamid. PhD. Dissertation committee. Goldberg Laboratory. University Program in Genetics and Genomics, Duke University. Now Scientist at Variant Bio.
- 2020-2021 Bobby Boone IV. Master's. Master's project committee. Landstrom Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now Biostatistician II at University of Utah Health.
- 2020-2023 Brandon M. Lê. PhD. Dissertation committee. Ashley-Koch Laboratory. University Program in Genetics and Genomics, Duke University.
- 2020-2023 Rachel Cason. Resident physician. Scholarship Oversight Committee, Master's project commit-



- tee. Gbadegesin Laboratory. Department of Pediatrics, Duke University.
- 2021-2022 Valerie Gartner. PhD. Dissertation committee. Wray Laboratory. University Program in Genetics and Genomics, Duke University.
- 2021-2022 Hongyu Du. Master's. Master's project committee. Allen Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- 2021-2022 Krista Pipho. PhD. Dissertation committee. Goldberg Laboratory. University Program in Genetics and Genomics, Duke University.
- 2021-2022 Weiliang "Frank" Tian. Master's. Master's project committee. Soderling Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- 2023-2023 Grace E. Rhodes. PhD. Graduate rotation. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University.

## Research experience

- 2013-2018 Postdoctoral Research, Storey Lab, Princeton. Developed new definitions and tools to study arbitrary population structures. Also applied GWAS and mixed effects modeling to longitudinal psychiatric drug clinical trials.
- 2007-2013 Graduate Research Thesis, Singh Lab and Llinás Lab, Princeton. Developed probabilistic models and statistical methods that improve protein domain prediction. Studied the *Plasmodium falciparum* proteome, focusing on the AP2 transcription factors. Used experimental techniques (cloning, protein purification and protein-binding microarrays) to test predicted AP2 domains for DNA binding.
- 2007 Graduate Rotation, Tavazoie Lab, Princeton. Cloned an RNA aptamer for a phage display assay.
- 2006-2007 Graduate Rotation, Singh Lab, Princeton. Analyzed protein interactions predicted from the presence of interacting domain pairs.
- 2006 Graduate Rotation, Troyanskaya Lab, Princeton. Built and analyzed a potential "gold standard" of Gene Ontology terms integrated from predictions of multiple sources, in *Saccharomyces cerevisiae* and *Homo sapiens*.
- 2005-2006 Undergraduate Research Assistant, Keating Lab, MIT. Analyzed full-atom computational protein designs using modified van der Waals potentials.

## Extracurricular University Activities

- 2006-2010 Latino Graduate Student Association, Princeton University. Board member: Technology Specialist 2007-2010 (handled website, email list, photos, calendar).
- 2002-2006 Association of Puerto Rican students, MIT. Regular member.

## Outreach

- 2017 Panelist at HISPA Latinos in College Conference. Princeton University.
- 2010 Ivy-plus recruiting fair at University of Puerto Rico (4 campuses) for Princeton.
- 2008-2011 Science and Engineering Expo. HHMI and Princeton University.
- 2007 Helped local high school students with college personal statements. PUPP, Princeton University.

## Languages

English. Native proficiency.

Spanish. Native proficiency.

French. Fluent reading and writing, conversational speaking.