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 https://ochoalab.github.io/

Education

2006

2008

2006

2001

2013-2018

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)

BS in Biology and Mathematics, Massachusetts Institute of Technology

Awards & Honors

Whitehead Scholar, Whitehead Charitable Foundation. For Duke junior faculty with exceptional potential for research and research training in the biomedical sciences.

NSF Graduate Research Fellowship. Computational Biology.

Ford Foundation Diversity Fellowship Predoc. Comp. (Declined, listed as Honorable Mention.)

MIT Department of Biology Merck Prize. Awarded to a graduating senior for outstanding research and academic performance in biophysics or bioinformatics.

Mexican Math Olympiad Gold Medal. Awarded to top 15 competitors nationally.

Employment

Assistant Professor, Duke Center for Statistical Genetics and Genomics, Department of Biostatistics and Bioinformatics, Duke University

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

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.

- Yiqi Yao, **Alejandro Ochoa**. Limitations of principal components in quantitative genetic association models for human studies. *eLife*. 79238. PMID 37140344.
- Zhuoran Hou, **Alejandro Ochoa**. Genetic association models are robust to common population kinship estimation biases. *Genetics*. iyado3o. PMID 36843304.
- 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 Gbadegesin. CLVS1, a candidate gene for steroid sensitive nephrotic syndrome, regulates podocyte oxidative stress and endocytosis. *JCI Insight*. 7(2) e152102. PMID 34874915.
- Brian I Shaw, **Alejandro Ochoa**, Cliburn Chan, Chloe Nobuhara, Rasheed Gbadegesin, 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.
- 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.
- Alejandro Ochoa, John D Storey. Estimating F_{ST} and kinship for arbitrary population structures. *PLoS Genet.* 17(1) e1009241. PMID 33465078.
- Alejandro Ochoa, Mona Singh. Domain prediction with probabilistic directional context. *Bioinf.* 33(16) 2471-8. PMID 28407137.
- Simon A Cobbold, Joana M Santos, **Alejandro Ochoa**, David H Perlman, Manuel Llinás. Proteomewide analysis reveals widespread lysine acetylation of major protein complexes in the malaria parasite. *Sci Rep.* 2016;6:19722. PMID 26813983.
- 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.
- 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.
- Alejandro Ochoa, Manuel Llinás, Mona Singh. Using context to improve protein domain identification. *BMC Bioinformatics*. 12:90. PMID 21453511.
- 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

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

ACKNOWLEDGMENTS

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.

- 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.
- 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

- simfam: Simulate and Model Family Pedigrees With Structured Founders. R, C++.
 - Available on CRAN and https://github.com/OchoaLab/simfam.
- 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.
- popkin: Estimate Kinship and FST under Arbitrary Population Structure. R, C++.
 - Available on CRAN and https://github.com/StoreyLab/popkin.
- bnpsd: Model and Simulate Admixed Populations. R.
 - Available on CRAN and https://github.com/StoreyLab/bnpsd.
- dPUC2: Domain Prediction Using Context, Version 2. Perl, C.
 - https://github.com/alexviiia/dpuc2.
- DomStratStats: Domain Stratified Statistics (q-values and local FDRs). Perl.
 - https://github.com/alexviiia/DomStratStats.
- 2014-2019 RandProt: High-order Markov random models for protein sequences. Perl.
 - https://github.com/alexviiia/RandProt.

MINOR PACKAGES

- simgenphen: Simulate Genotypes and Phenotypes. R.
 - https://github.com/OchoaLab/simgenphen.
- 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.
- popkinsuppl: Supplement to "popkin" package. R.
 - https://github.com/OchoaLab/popkinsuppl.

Paper repositories

- 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.
- data: Instructions for real data processing shared across projects. R, bash, Perl, markdown.
 - https://github.com/OchoaLab/data.

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

Invited Talks

2019

- HCHS/SOL (Hispanic Community Health Study / Study of Latinos) Genetics Special Interest Group (Virtual). Department of Biostatistics, Harvard University, Cambridge, MA. 2023-06-21.
- NIH Centers for Excellence in Genomics Science (CEGS) Virtual Outreach Meeting. The Jackson Laboratory, Bar Harbor, ME. 2023-04-20.
- American Mathematical Society (AMS) 2023 Spring Southeastern Sectional Meeting. Georgia Institute of Technology, Atlanta, GA. 2023-03-19.
- BERD Core Seminar. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2022-08-26.
- 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.
- Duke University Program in Genetics and Genomics (UPGG) retreat. Suite 4, Durham, NC. 2022-06-11.
- Orientation for MB and PhD students. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2021-08-19.
- Statistics seminar. Department of Mathematics and Statistics, Washington State University, Pullman, WA. 2021-03-10.
- Master of Biostatistics Virtual Visit Day. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2021-02-19.
- University Program in Genetics and Genomics (UPGG) seminar. Room 147, Nanaline Duke building, Duke University, Durham, NC. 2020-02-18.
- Duke Computational Biology and Bioinformatics (CBB) retreat. Holiday Inn Resort, Wrightsville Beach, NC. 2019-09-28.
- Computational Biology and Bioinformatics (CBB) seminar. French Family Science Center, Duke University, Durham, NC. 2019-02-04.
- *4th Mexico Population Genomics Meeting.* Amoxcalli Complex, Faculty of Sciences, University City, National Autonomous University of Mexico, Mexico City, DF, Mexico. 2019-01-11.
- Population Biology seminar, Department of Biology, Duke University, Durham, NC. 2018-09-27.
- Department of Biostatistics and Bioinformatics, Duke University, Durham, NC. 2018-02-21.
- Department of Biostatistics, Johns Hopkins University, Baltimore, MD. 2018-01-22.
- Department of Biology, University of Richmond, Richmond, VA. 2017-11-10.
- Princeton Research Day. Princeton University, Princeton, NJ. 2017-05-11.
- Department of Genetics, University of North Carolina, Chapel Hill, NC. 2017-02-13.
- New York Area Population Genomics Workshop 2016. Princeton University, Princeton, NJ. 2016-01-21.
- 2015 Probabilistic Modeling in Genomics conference. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 2015-10-15.
- Telepresentation for Yun Song's group. UC Berkeley, Berkeley, CA. 2013-04-17.
- Biological sequence analysis and probabilistic models conference. HHMI Janelia Farm, Ashburn, VA. 2013-03-25.

- NCBI, NIH, Bethesda, MD. 2013-02-25.
- Recruiting conference. Department of Computer Science, Princeton University, Princeton, NJ. 2012-03-01.
- Rotary Club Paso del Norte, Ciudad Juarez, CHIH, Mexico. 2004-08-26.

Conference posters

- 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.
- Alejandro Ochoa, Yuncheng Duan, Revathy Venukuttan, Shannon Clarke, Timothy 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.
- 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.
- 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.
- Alejandro Ochoa, John D Storey. Relatedness and Differentiation in Arbitrary Population Structures. *Probabilistic Modeling in Genomics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Alejandro Ochoa, John D Storey. Relatedness and Differentiation in Arbitrary Population Structures. *Population, Evolutionary and Quantitative Genetics Conference*. Madison Concourse Hotel, Madison, WI.
- 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.
- 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.
- 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.
- 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 conference*. HHMI Janelia Farm, Ashburn, VA.
- Alejandro Ochoa, Manuel Llinás, Mona Singh. Using context to predict protein domains across diverse organisms. *Recomb Systems Biology conference*. Columbia University, New York, NY.

Teaching

Dire	LIMINEDCITY
LJUKE	UNIVERSITY

	DUKE UNIVERSITY
2022-2023 2021-2023 2019-2023 2019 2019	Human Genetics. (Co-Lecturer Spring 2022-2023) UPGEN 778A-F Mini course: Genetic population structure and relatedness (Fall 2021-2023) SIBS lecture on genetic association studies (Summer 2019, 2021-2023) BIOS 900 Current Problems in Biostatistics. Special lecture on kinship and $F_{\rm ST}$ BIOS 710 Statistical Genetics and Genetic Epidemiology. Special lecture on kinship and $F_{\rm ST}$
	Princeton University
2017 2016-2017 2014-2017 2011 2008	Intro to Genomics and Comp Bio. Quant Comp Bio, Comp Sci. (Co-Lecturer Fall 2017) Discussion Leader. Summer Undergraduate Research Program (Summer 2016, 2017) Statistical Treatment of Data. Mol Bio, Quant Comp Bio. (Workshop Fall 2014-2017) Intro to Genomics and Comp Mol Bio. Mol Bio, Comp Sci. (TA Fall 2011) Core Laboratory. Mol Bio. (TA Spring 2008)
	Massachusetts Institute of Technology
2005 2003-2004 2003-2004	Calculus. OME Project Interphase (TA Summer 2005) Multivar Calculus and Diff Eqs. OME Seminar XL (Tutor Fall 2003, Spring 2004) Calculus. MITE2S Program (TA Summer 2003, 2004)
	Mentoring
	Mentoring Current Trainees - Primary Adviser
2020-now 2022-now 2022-now 2023-now	
2022-now 2022-now	Current Trainees - Primary Adviser Tiffany Tu. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. Computational Biology and Bioinformatics Program, Duke University. Ratchanon "RP" Pornmongkolsuk. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. University Program in Genetics and Genomics, Duke University. Zhuoran Hou. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Danielle Mensah. Undergraduate. Research Independent Study. Ochoa Laboratory. Department
2022-now 2022-now	Current Trainees - Primary Adviser Tiffany Tu. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. Computational Biology and Bioinformatics Program, Duke University. Ratchanon "RP" Pornmongkolsuk. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. University Program in Genetics and Genomics, Duke University. Zhuoran Hou. PhD. Graduate rotation, Dissertation adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Danielle Mensah. Undergraduate. Research Independent Study. Ochoa Laboratory. Department of Computer Science, Duke University.

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

PAST TRAINEES - PRIMARY ADVISER

- 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.
- 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.
- Zhuoran Hou. Master's. BCTIP internship. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now PhD Student at Duke University, BB.
- Jiajie Shen. Master's. Research, Master's project adviser. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- Emmanuel Mokel. Undergraduate. Research Independent Study. Ochoa Laboratory. Department of Statistical Science, Duke University.

PAST TRAINEES - SECONDARY ADVISER

- Neo Christopher Chung. PhD. Graduate rotation. Llinás Laboratory. Quantitative Computational Biology Program, Princeton University. Now Adjunct Faculty at University of Warsaw.
- Jeremy Bigness. PhD. Graduate rotation. Singh Laboratory. Quantitative Computational Biology Program, Princeton University. Now PhD Student at Brown University.
- 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.
- Yuncheng Duan. PhD. Graduate rotation. Ochoa Laboratory. Department of Biology, Duke University. Now PhD Student at Duke University, Allen Lab.
- Shengyu Li. Master's. Master's project committee. Allen Laboratory. Department of Biostatistics and Bioinformatics, Duke University. Now PhD Student at Duke University, CBB.
- Iman Hamid. PhD. Dissertation committee. Goldberg Laboratory. University Program in Genetics and Genomics, Duke University. Now Scientist at Variant Bio.
- 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.
- 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.
- Valerie Gartner. PhD. Dissertation committee. Wray Laboratory. University Program in Genetics and Genomics, Duke University.
- Hongyu Du. Master's. Master's project committee. Allen Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- Krista Pipho. PhD. Dissertation committee. Goldberg Laboratory. University Program in Genetics and Genomics, Duke University.
- ²⁰²¹⁻²⁰²² Weiliang "Frank" Tian. Master's project committee. Soderling Laboratory. Department of Biostatistics and Bioinformatics, Duke University.
- Grace E. Rhodes. PhD. Graduate rotation. Ochoa Laboratory. Department of Biostatistics and Bioinformatics, Duke University.

Research experience

- 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.
- 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.
- Graduate Rotation, Tavazoie Lab, Princeton. Cloned an RNA aptamer for a phage display assay.

 Graduate Rotation, Singh Lab, Princeton. Analyzed protein interactions predicted from the presence of interacting domain pairs.
- 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*.
- Undergraduate Research Assistant, Keating Lab, MIT. Analyzed full-atom computational protein designs using modified van der Waals potentials.

Extracurricular University Activities

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

Outreach

- Panelist at HISPA Latinos in College Conference. Princeton University.
- Ivy-plus recruiting fair at University of Puerto Rico (4 campuses) for Princeton.
- Science and Engineering Expo. HHMI and Princeton University.
- 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.

Last updated: February 22, 2024