Leonardo Collado-Torres

855 N. Wolfe Street, Suite 300 – 21205 – United States

+1 (410) 955 1000 • lcolladotor@gmail.com
lcolladotor.github.io • fellgernon • lcolladotor

Education

Johns Hopkins Bloomberg School of Public Health — **Baltimore, US**
PhD in Biostatistics — *2011-2016*

National Autonomous University of Mexico (UNAM) — **Cuernavaca, MX**
Bachelor in Genomic Sciences (LCG), Grade 9.71/10 — *2005-2009*

ITESM Campus Cuernavaca — **Cuernavaca, MX**
High school, Grade 97.8/100 — *2002-2005*

PhD thesis

Title: *Annotation-Agnostic Differential Expression and Binding Analyses*.

Advisors: **Jeffrey T. Leek** and **Andrew E. Jaffe**.

Description: The goal was to develop statistical methods and software that enable researchers to differentiate the sources of variation observed in RNA-seq while minimizing the dependance on known annotation. This allows researchers to correct for technological variation and study the biological variation driving their phenotype of interest. We applied these methods to further our understanding of neuropsychiatric disorders using the Lieber Institute for Brain Development human brains collection ($>$ 1000 samples).

Honors and awards

2019: Travel award for BioC2019.

2019: rstudio::conf 2019 diversity award recipient RStudio.

2019: Early Career Clinical Research Symbiont Award researchsymbionts.org.

2018: National Researcher level 1, CONACyT Mexico (results): *Investigador Nacional Nivel 1*.

2018: rOpenSci Unconf 2018: unconf18 participants and application.

2018: Winner entry for the Bioinformatics Peer Prize III.

2017: Scholarship to attend and present at BioC2017.

2014: Scholarship to attend BioC2014.

2013: Travel award winner for useR2013.

2011: Scholarship to attend BioC2011.

2011: Awarded CONACyT Mexico scholarship for PhD studies outside Mexico.

2010: Travel award to attend From Functional Genomics to Systems Biology 2010 and BiocDevelEurope 2010.

2009: *Summa Cum Laude* for bachelor in Genomic Sciences studies at LCG-UNAM.

2005: Best high school average ($\sim$ 200 students): awarded ITESM system 90% scholarship for college studies, declined to join LCG-UNAM.

Experience

Research

Andrew Jaffe lab **LIBD, Baltimore, US**

Staff Scientist *2016–current*

At LIBD, Dr. Collado-Torres is part of the Data Science team which goals include better understanding and characterizing genomics signatures in the human brain, including DNA methylation and gene expression. Leonardo helps mentor other team members, provides support for LIBD projects and is advancing his academic career as part of Andrew Jaffe's lab. He is also a co-founder of the LIBD rstats club and CDSBMexico.

Enrique Morett lab **IBT-UNAM, Cuernavaca, MX**

Bioinformatician *2009–2011*

Identified transcriptions start sites and transcription units in *Escherichia coli* and *Geobacter sulfurreducens* with RNA-seq data. Developed the *BacterialTranscription* R package.

Guillermo Dávila lab **CCG-UNAM, Cuernavaca, MX**

Undergraduate research assistant *2007–2009*

Determined bacteriophage ecological groups by developing a method based on codon distribution of all phage sequenced genomes. Joint work with Sur Herrera Paredes.

Roberto Kolter lab **Harvard, Boston, US**

Undergraduate research assistant *2007*

Supervisor: Elizabeth Shank. Carried out screenings to identify bacteria that activate the production of exopolysaccharide through the activation of the gene tasA in *Bacillus subtilis*.

Industry

Winter Genomics **Cuernavaca, MX**

Data Science Division Leader *2009-2011*

Responsible for recruiting and hiring new personnel, overseeing and supervising bioinformaticians, training new employees, writing research reports and presenting them to colleagues, and organizing all scientific projects.

- First scientific staff member at Winter Genomics;
- Projects completed:
  - de novo genome assembly simulations,
  - assembly and annotation of the *phiVC8* bacteriophage genome,
  - integrated analysis of more than 20 RNA-seq samples for determination of transcription initiation in *Escherichia coli* reported in Gama-Castro et al., PMID 21051347,
  - de novo assembly of four *Escherichia coli* strains and lead to Aguilar et al., PMID 22884033;
- Designed training material for new employees.

Publications

∗ indicates equal contribution, † indicates corresponding author

Pre-prints

1. Eddie-Luidy Imada*, Diego Fernando Sanchez*, **Leonardo Collado-Torres**, Christopher Wilks, Tejasvi Matam, Wikum Dinalankara, Aleksey Stupnikov, Francisco Lobo-Pereira, Chi-Wai Yip, Kayoko Yasuzawa, Naoto Kondo, Masayoshi Itoh, Harukazu Suzuki, Takeya Kasukawa, Chung-Chau Hon, Michiel JL de Hoon, Jay W Shin, Piero Carninci, FANTOM consortium, Andrew E Jaffe, Jeffrey T Leek, Alexander Favorov, Gloria R Franco, Ben Langmead†, Luigi Marchionni†. Recounting the FANTOM Cage Associated Transcriptome. *bioRxiv* 659490 (2019). doi: 10.1101/659490.
2. Ashkaun Razmara, Shannon E Ellis, Dustin J Sokolowski, Sean Davis, Michael D Wilson, Jeffrey T Leek, Andrew E Jaffe, **Leonardo Collado-Torres**†. *recount-brain*: a curated repository of human brain RNA-seq datasets metadata. *bioRxiv* 618025 (2019). doi: 10.1101/618025.
3. Sebastian Guelfi*, Karishma D'Sa*, Juan Botía*, Jana Vandrovcova, Regina H. Reynolds, David Zhang, Daniah Trabzuni, **Leonardo Collado-Torres**, Andrew Thomason, Pedro Quijada Leyton, Sarah A. Gagliano, Mike A. Nalls, UK Brain Expression Consortium, Kerrin S. Small, Colin Smith, Adaikalavan Ramasamy, John Hardy, Michael E. Weale†, Mina Ryten†. Regulatory sites for known and novel splicing in human basal ganglia are enriched for disease-relevant information. *bioRxiv* 591156 (2019). doi: 10.1101/591156.
4. David Zhang*, Sebastian Guelfi*, Sonia Garcia Ruiz, Beatrice Costa, Regina H Reynolds, Karishma D'Sa, Wenfei Liu, Thomas Courtin, Amy Peterson, Andrew E Jaffe, John Hardy, Juan Botia, **Leonardo Collado-Torres**, Mina Ryten. Incomplete annotation of OMIM genes is likely to be limiting the diagnostic yield of genetic testing, particularly for neurogenetic disorders. *bioRxiv* 499103 (2018). doi: 10.1101/499103.
5. Carrie Wright*, Anandita Rajpurohit*, Emily E. Burke, Courtney Williams, **Leonardo Collado-Torres**, Martha Kimos, Nicholas J. Brandon, Alan J. Cross, Andrew E. Jaffe, Daniel R. Weinberger†, Joo Heon Shin†. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *bioRxiv* 445437 (2018). doi: 10.1101/445437.
6. Amanda J. Price*, **Leonardo Collado-Torres***, Nikolay A. Ivanov, Wei Xia, Emily E. Burke, Joo Heon Shin, Ran Tao, Liang Ma, Yankai Jia, Thomas M. Hyde, Joel E. Kleinman, Daniel R. Weinberger, Andrew E Jaffe. Divergent neuronal DNA methylation patterns across human cortical development: Critical periods and a unique role of CpH methylation. *bioRxiv* 428391 (2018). doi: 10.1101/428391.
7. Emily E Burke*, Joshua G Chenoweth*, Joo Heon Shin, **Leonardo Collado-Torres**, Suel Kee Kim, Nicola Micali, Yanhong Wang, Richard E Straub, Daniel J Hoeppner, Huei-Ying Chen, Alana Lescure, Kamel Shibbani, Gregory R Hamersky, BaDoi N Phan, William S Ulrich, Cristian Valencia, Amritha Jaishankar, Amanda J Price, Anandita Rajpurohit, Stephen A Semick, Roland Bürli, James C Barrow, Daniel J Hiler, Stephanie Cerceo Page, Keri Martinowich, Thomas M Hyde, Joel E Kleinman, Karen F Berman, José A Apud, Alan J Cross, Nick J Brandon, Daniel R Weinberger, Brady J Maher, Ronald DG McKay†,

Andrew E Jaffe†. Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs. *bioRxiv* 380758 (2018). doi: 10.1101/380758.

8. Fu J, Kammers K, Nellore A, **Collado-Torres L**, Leek JT, Taub MA. RNA-seq transcript quantification from reduced-representation data in recount2. *bioRxiv* 247346 (2018). doi: 10.1101/247346.

Peer-reviewed

1. **Leonardo Collado-Torres**, Emily E Burke, Amy Peterson, JooHeon Shin, Richard E Straub, Anandita Rajpurohit, Stephen A Semick, William S Ulrich, BrainSeq Consortium, Amanda J Price, Cristian Valencia, Ran Tao, Amy Deep-Soboslay, Thomas M Hyde, Joel E Kleinman, Daniel R Weinberger†, Andrew E Jaffe†. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. *Neuron* 2019. doi: 10.1016/j.neuron.2019.05.013
   Pre-print: *bioRxiv* 426213 (2018). doi: 10.1101/426213.
2. Stephen A Semick, Rahul A Bharadwaj, **Leonardo Collado-Torres**, Ran Tao, Joo Heon Shin, Amy Deep-Soboslay, James R. Weiss, Daniel R Weinberger, Thomas M Hyde, Joel E Kleinman, Andrew E Jaffe†, Venkata S Mattay†. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathologica* 2019. doi: 10.1007/s00401-019-01966-5
   Pre-print: *bioRxiv* 430603 (2018). doi: 10.1101/430603.
3. Helena Kuri-Magaña, **Leonardo Collado-Torres**, Andrew E Jaffe, Humberto Valdovinos-Torres, Marbella Ovilla-Muñoz, Juan M Téllez-Sosa, Laura C Bonifaz-Alfonzo, Jesús Martínez-Barnetche. Non-coding Class Switch Recombination-Related Transcription in Human Normal and Pathological Immune Responses. *Frontiers in Immunology* 2018. doi: 10.3389/fimmu.2018.02679
   Pre-print: *bioRxiv* 384172 (2018). doi: 10.1101/384172.
4. Semick SA, **Collado-Torres L**, Markunas CA, Shin JH, Deep-Soboslay A, Tao R, Huestis MA, Bierut LJ, Maher BS, Johnson EO, Hyde TM, Weinberger DR, Hancock DB, Kleinman JE†, Jaffe AE†. Developmental effects of maternal smoking during pregnancy on the human frontal cortex transcriptome. *Molecular Psychiatry* 2018. doi: 110.1038/s41380-018-0223-1
   Pre-print: *bioRxiv* 236968 (2017). doi: 10.1101/236968.
5. Jaffe AE, Straub R, Shin JH, Tao R, Gao Y, **Collado-Torres L**, Kam-Thong T, Xi HS, Quan J, Chen Q, Colantuoni C, Ulrich WS, Maher BJ, Deep-Soboslay A, The BrainSeq Consortium, Cross AJ, Brandon NJ, Leek JT, Hyde TM, Kleinman JE, Weinberger DR. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* 2018. doi: 10.1038/s41593-018-0197-y.
   Pre-print: *bioRxiv* 124321 (2017). doi: 10.1101/145656.
6. Ellis SE, **Collado-Torres L**, Jaffe AE, Leek JT. Improving the value of public RNA-seq expression data by phenotype prediction. *Nucl. Acids Res.* 2018. doi: 10.1093/nar/gky102.
   Pre-print: *bioRxiv* 145656 (2017). doi: 10.1101/145656.
7. **Collado-Torres L**†, Nellore A, Jaffe AE. recount workflow: Accessing over 70,000 human RNA-seq samples with Bioconductor [version 1; referees: 1 approved, 2 approved

with reservations]. *F1000Research* (2017). doi: 10.12688/f1000research.12223.1.
Winning entry for the Bioinformatics Peer Prize III.

8. Wright C, Shin JH, Rajpurohit A, Deep-Soboslay A, **Collado-Torres L**, Brandon NJ, Hyde TM, Kleinman JE, Jaffe AE, Cross AJ, Weinberger DR. Altered expression of histamine signaling genes in autism spectrum disorder. *Translational Psychiatry* 2017. doi: 10.1038/tp.2017.87.
9. **Collado-Torres L***, Nellore A*, Kammers K, Ellis SE, Taub MA, Hansen KD, Jaffe AE, Langmead B, Leek JT. Reproducible RNA-seq analysis using *recount2*. *Nature Biotechnology* 2017. doi: 10.1038/nbt.3838.
Pre-print: *bioRxiv* 068478 (2016). doi: 10.1101/068478.
10. Nellore A, Jaffe AE, Fortin JP, Alquicira-Hernández J, **Collado-Torres L**, Wang S, Phillips RA, Karbhari N, Hansen KD, Langmead B, Leek JT. Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. *Genome Biology* 2016. doi: 10.1186/s13059-016-1118-6.
Pre-print: *bioRxiv* 038224 (2016). doi: 10.1101/038224.
11. **Collado-Torres L**, Nellore A, Frazee AC, Wilks C, Love MI, Langmead B, Irizarry RA, Leek JT, Jaffe AE. Flexible expressed region analysis for RNA-seq with derfinder. *Nucl. Acids Res.* 2016. doi: 10.1093/nar/gkw852.
Pre-print: *bioRxiv* 015370 (2016). doi: 10.1101/015370.
12. Nellore A, **Collado-Torres L**, Jaffe AE, Alquicira-Hernández J, Wilks C, Pritt J, Morton J, Leek JT, Langmead B. Rail-RNA: Scalable analysis of RNA-seq splicing and coverage. *Bioinformatics* 2016. doi: 10.1093/bioinformatics/btw575.
Pre-print: *bioRxiv* 019067 (2015). doi: 10.1101/019067.
13. **Collado-Torres L**, Jaffe AE and Leek JT. regionReport: Interactive reports for region-level and feature-level genomic analyses [version2; referees: 2 approved, 1 approved with reservations]. *F1000Research* 2016, 4:105. doi: 10.12688/f1000research.6379.2.
Pre-print: *bioRxiv* 016659 (2015). doi: 10.1101/016659.
14. Jaffe AE, Shin J, **Collado-Torres L**, Leek JT, et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat. Neurosci.* 2015. doi: 10.1038/nn.3898.
15. Shank EA, Klepac-Ceraj V, **Collado-Torres L**, Powers GE, Losick R, Kolter R. Interspecies interactions that result in Bacillus subtilis forming biofilms are mediated mainly by members of its own genus. *Proc. Natl. Acad. Sci.* U.S.A. 2011 Nov;108(48):E1236–1243. doi: 10.1073/pnas.1103630108.
16. Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muñiz-Rascado L, Solano-Lira H, Jimenez-Jacinto V, Weiss V, García-Sotelo JS, López-Fuentes A, Porrón-Sotelo L, Alquicira-Hernández S, Medina-Rivera A, Martínez-Flores I, Alquicira-Hernández K, Martínez-Adame R, Bonavides-Martínez C, Miranda-Ríos J, Huerta AM, Mendoza-Vargas A, **Collado-Torres L**, Taboada B, Vega-Alvarado L, Olvera M, Olvera L, Grande R, Morett E, Collado-Vides J. RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* 2011 Jan;39(Database issue):D98–105. doi: 10.1093/nar/gkq1110.

Books

1. Frazee AC, **Collado-Torres L**, Jaffe AE, Langmead B, Leek JT. Measurement, Summary, and Methodological Variation in RNA-sequencing in Statistical Analysis of Next Generation Sequencing Data, *Springer*, 2014, 115-128.

Public profiles

Google Scholar: h57-MykAAAAJ

ORCID: 0000-0003-2140-308X

Impactstory: 0000-0003-2140-308X

GitHub: lcolladotor

Twitter: fellgernon

Speaker Deck: lcolladotor

SlideShare: lcolladotor

LinkedIn: lcollado

Epernicus: lc40

publons: 1262671

Professional service

Develop and maintain open-source biostatistical software.

Peer review

Statistical Applications in Genetics and Molecular Biology: Since 2017

F1000Research: Collado-Torres L. Referee Report For: BgeeDB, an R package for retrieval of curated expression datasets and for gene list expression localization enrichment tests [version 1; referees: 1 approved, 1 approved with reservations, 1 not approved]. F1000Research 2016, 5:2748 (doi: 10.5256/f1000research.10748.r17980)

Bioinformatics: Since 2015

Biostatistics: Since 2013

Professional memberships

2019: International Society for Computational Biology

2015,2018: American Society of Human Genetics

2015-2017: American Statistical Association

2014-2016: ENAR student member

2014: American Public Health Association

Presentations

Most recent slides are available via speakerdeck or slideshare.

Talks at conferences

2019: Reproducible RNA-seq analysis with recount2, *PSB2019*, Hawaii – US. (slides)

2018: *Keynote* From learning to using to teaching to developing R, *CDSBMexico*, Cuernavaca – MX (remote presentation). (slides)

2018: Unique Molecular Correlates of Schizophrenia and Its Genetic Risk in the Hippocampus Compared to Frontal Cortex, *SOBP*, New York – US. (slides)

2018: BrainSeq Phase II: Schizophrenia-associated expression differences between the hippocampus and the dorsolateral prefrontal cortex, *BOG*, Cold Spring Harbor – US. (slides)

2018: recount workflow: Accessing over 70,000 human RNA-seq samples with Bioconductor, *Online Journal Club by Dennis Lal*, Online. (slides)

2017: Guiding principles for interactive graphics based on LIBD data science projects, *JSM*, Baltimore – US. (slides)

2017: Reproducible RNA-seq analysis with recount2 workshop, *BioC*, Boston – US. (slides)

2017: Reproducible RNA-seq analysis with recount2, *ICSA*, Chicago – US. (slides)

2017: RNA-seq samples beyond the known transcriptome with derfinder and recount, *SOBP*, San Diego – US. (slides)

2016: recount: facilitando el análisis de miles de muestras de RNA-seq, *Genomeeting2016*, Mexico City – MX. (slides)

2016: Using Data Science to Study Human Brain Genomic Measurements, *SACNAS*, Long Beach – US. (slides)

2016: **Collado-Torres L**, et al. Annotation-agnostic differential expression analysis, *ENAR*, Austin – US. (slides)

2015: **Collado-Torres L**, Frazee AC, Love MI, Irizarry RA, Jaffe AE, Leek JT. Annotation-agnostic differential expression analysis, *Genomics and Bioinformatics Symposium*, Center for Computational Genomics, Hopkins, Baltimore – US. (slides)

2015: Jaffe AE, Shin J, **Collado-Torres L**, Leek JT, et al. Dissecting human brain development at high resolution using RNA-seq, *ENAR*, Miami – US. (slides)

2014: Jaffe AE, Shin J, **Collado-Torres L**, Leek JT, et al. Developmental regulation of human cortex transcription at base-pair resolution, *is3b*: 1st International Summer Symposium on Systems Biology, INMEGEN, Mexico City – MX. (slides)

2014: **Collado-Torres L**, Frazee AC, Love MI, Irizarry RA, Jaffe AE, Leek JT. Fast differential expression analysis annotation-agnostic across groups with biological replicates, LCG 10 year anniversary, LCG-UNAM, Cuernavaca – MX. (slides)

2013: **Collado-Torres L**, Frazee AC, Irizarry RA, Jaffe AE, Leek JT. Differential expression analysis of RNA-seq data at base-pair resolution in multiple biological replicates, *useR2013*, Albacete – Spain. (slides)

2010: **Collado-Torres L**, Reyes-Quiroz A, Cuéllar-Partida G, Moreno-Mayar V, Vargas-Chávez C, Collado-Vides J. BacterialTranscription: a R package to identify Transcription

Start Sites and Transcription Units, *Bioconductor Developer Meeting*, EMBL, Heidelberg – Germany. (slides)

Posters

2019: *recount-brain*: a curated repository of human brain RNA-seq datasets metadata, *BoG2019*, Cold Spring Harbor – US. (PDF)

2018: Regional heterogeneity in gene expression, regulation and coherence in hippocampus and dorsolateral prefrontal cortex across development and in schizophrenia, *ASHG2018*, *biodata18* and *PSB2019*, San Diego, Cold Spring Harbor and Hawaii – US. (PDF)

2018: BrainSeq Phase II: schizophrenia-associated expression differences between the hippocampus and the dorsolateral prefrontal cortex, *SAGES2018*, Philadelphia – US. (PDF)

2017: **Collado-Torres L**†, Nellore A, Jaffe AE. Getting started with recount2 and accessing it via R, *IDIES2017*, Baltimore – US. (PDF)

2015: **Collado-Torres L**, Frazee AC, Love MI, Irizarry RA, Jaffe AE, Leek JT. Annotation-agnostic RNA-seq differential expression analysis software, *ASHG2015* and *IDIES2015*, Baltimore – US. (PDF)

2014: **Collado-Torres L**, Frazee AC, Love MI, Irizarry RA, Jaffe AE, Leek JT. Fast annotation-agnostic differential expression analysis, *ENAR* and *Delta Omega Poster Competition (JHBSPH)*, Baltimore – US. (PDF)

2013: **Collado-Torres L**, Jaffe AE, Leek JT. Fast annotation-agnostic differential expression analysis, *Genomics and Bioinformatics Symposium*, Center for Computational Genomics, Hopkins, Baltimore – US. (PDF)

2013: **Collado-Torres L**, et al. Differential expression RNA-seq analysis with a large data set from brain samples, *JHU Biostatistics Department Retreat*, Philadelphia – US. (PDF)

2010: **Collado-Torres L**, Reyes-Quiroz A, Cuéllar-Partida A, Moreno-Mayar V, Taboada B, Vega-Alvarado L, Jiménez-Jacinto V, Mendoza-Vargas A, Grande R, Olvera L, Olvera M, Vargas-Chávez C, Júarez K, Collado-Vides J, Morett E. Global Analysis of Transcription Start Sites and Transcription Units in Bacterial Genomes, *From Functional Genomics to Systems Biology*, EMBL, Heidelberg – Germany. (PDF)

2010: **Collado-Torres L**, Reyes-Quiroz A, Cuéllar-Partida A, Moreno-Mayar V, Taboada B, Vega-Alvarado L, Jiménez-Jacinto V, Mendoza-Vargas A, Grande R, Olvera L, Olvera M, Vargas-Chávez C, Júarez K, Collado-Vides J, Morett E. Global Analysis of Transcription Start Sites and Transcription Units in Bacterial Genomes, *BioC2010*, FHCRC, Seattle – US. (PDF)

Other talks

2019: Analyzing BrainSeq Phase II and generating the recount-brain resource, *Staff Seminar Series*, LIBD, Baltimore – US. (slides)

2018: recount-brain: a curated repository of human brain RNA-seq datasets metadata, *Joint Genomic Meeting*, JHU, Baltimore – US. (slides)

2018: Reproducible RNA-seq analysis with recount2 and recount-brain, LCG-UNAM via Skype. (slides)

2017: Reproducible Research and Bioinformatics, *Summer Institute*, JHBSPH, Baltimore – US. (slides)

2016: Introduction at Kandahar University MPH training event. (slides)

2015: dbFinder, *Joint Genomic Meeting*, JHBSPH, Baltimore – US. (slides)

2015: Easy parallel computing with BiocParallel and HTML reports with knitrBootstrap, *Biostatistics Computing Club*, JHBSPH, Baltimore – US. (slides)

2015: Does mapping simulated RNA-seq reads provide information?, *Joint Genomic Meeting*, JHBSPH, Baltimore – US. (slides)

2014: derfinder tutorial, *Leek group lab meeting*, JHBSPH, Baltimore – US. (slides)

2014: Git for research, *Biostatistics Computing Club*, JHBSPH, Baltimore – US. (slides)

2015: Does mapping simulated RNA-seq reads provide information?, *Joint Genomic Meeting*, JHBSPH, Baltimore – US. (slides)

2013: Introduction to ggbio, *Genomics for Students*, JHBSPH, Baltimore – US. (slides)

2013: Fast differential expression analysis annotation-agnostic across groups with biological replicates, *Joint Genomic Meeting*, JHBSPH, Baltimore – US. (slides)

2013: Fast differential expression analysis annotation-agnostic across groups with biological replicates, *Biostatistics Journal Club*, JHBSPH, Baltimore – US. (slides)

2013: Introduction to knitr, *Biostatistics Computing Club*, JHBSPH, Baltimore – US. (slides)

2013: Introduction to High-Throughput Sequencing and RNA-seq, *Genomics for Students*, JHBSPH, Baltimore – US. (slides)

2012: DEXSeq paper discussion, *Genomics for Students*, JHBSPH, Baltimore – US. (slides)

2012: Introduction to R and Biostatistics, LCG-UNAM via Skype. (slides)

2012: Introducing Git while making your academic webpage, *Biostatistics Computing Club*, JHBSPH, Baltimore – US. (slides)

2011: Introducing Biostatistics to first year LCG students, LCG-UNAM via Skype. (slides)

2010: Introduction to using Bioconductor for High Throughput Sequencing Analysis, *National Bioinformatics Week*, CCG-UNAM, Cuernavaca – MX. (slides)

2009: Bacteriophages: analyzing their diversity, *LCG third generation symposium*, CCG-UNAM, Cuernavaca – MX. (slides)

Courses and Meetings Attendance

2019: *BoG*, Cold Spring Harbor – US.

2019: *RStudio conf*, Austin – US.

2019: *Pacific Symposium on Biocomputing (PSB)*, Hawaii – US.

2018: *Biological Data Science (biodata18)*, Cold Spring Harbor – US.

2018: *ASHG*, San Diego – US.

2018: *SAGES*, Philadelphia – US.

2018: *SOBP*, New York – US.

2018: *BoG*, Cold Spring Harbor – US.

2018: *rOpenSci unconf*, Seattle – US.

2017: *JSM*, Baltimore – US.

2017: *BioC*, Dana-Farber Cancer Institute, Boston – US.

2017: *ICSA*, Chicago – US.

2017: *SOBP*, San Diego – US.

2016: *SACNAS*, Long Beach – US.

2016: *ENAR*, Austin – US.

2015: *ENAR*, Miami – US.

2014: *is3b*, INMEGEN, Mexico City – MX.

2014: *BioC*, Dana-Farber Cancer Institute, Boston – US.

2014: *IDIES2014*, Johns Hopkins University, Baltimore – US.

2014: *ENAR*, Baltimore – US.

2014: *Delta Omega Poster Competition*, Johns Hopkins University, Baltimore – US.

2014: LCG 10 year anniversary, LCG-UNAM, Cuernavaca – MX.

2013: *Genomics and Bioinformatics Symposium*, Johns Hopkins University, Baltimore – US.

2013: *useR2013*, Albacete – Spain.

2011: *BioC*, FHCRC, Seattle – US.

2010: *From Functional Genomics to Systems Biology*, EMBL, Heidelberg – Germany.

2010: *BioC*, FHCRC, Seattle – US.

2009: *BioC*, FHCRC, Seattle – US.

2009: Course on Oral Communication taught by the master Rafael Popoca, CCG-UNAM, Cuernavaca – MX.

2008: *BioC*, FHCRC, Seattle – US.

2008: *A Short R/Bioconductor Course* by James Bullard from UC Berkeley, LCG-UNAM, Cuernavaca – MX.

2007: *Boston Bacterial Meeting*, Boston – US.

2007: *Retreat of the Department of Microbiology and Molecular Genetics - Harvard*, Boston – US.

2006: *Winter School in Genomics*, CCG-UNAM, Cuernavaca – MX.

2005: *HUGO 2005*, Kyoto – Japan.

Software

Bioconductor – main author

2016: recount: Explore and download data from the recount project – 15812 downloads.

2014: derfinder: Annotation-agnostic differential expression analysis of RNA-seq data at base-pair resolution via the DER Finder approach – 23968 downloads.

2014: derfinderPlot: plotting functions for derfinder results – 10992 downloads.

2014: regionReport: Generate HTML or PDF reports for a set of genomic regions or DESeq2/edgeR results – 12426 downloads.

2014: derfinderHelper: helper functions for derfinder package – 19809 downloads.

2014: derfinderData: data for derfinder examples – 2515 downloads.

Bioconductor – contributor role

2015: bumphunter

2014: ballgown

Other R packages

2018: blogdown: contributor role, mainly in the *Insert Image* and *New Post* addins.

2017: recount.bwtool: Compute coverage matrices from recount quickly using bwtool.

2016-2018: jaffelab: commonly used functions by the Jaffe lab.

2016: shinycsv: explore a table interactively.

2014: enrichedRanges: identify enrichment between two sets of genomic ranges.

2014: dots: simplify function calls.

2013: fitbitR: visualize your FitBit data.

2011: BacterialTranscription: identify TSSs and TUs from RNA-seq data.

shiny applications

2016: shinycsv: explore a table interactively deployed at jhubiostatistics.shinyapps.io.

2016: recount: analysis-ready RNA-seq gene and exon counts datasets deployed at jhubiostatistics.shinyapps.io.

2014–2016: MPH capstone TA office hours sign up.

2014: Simple mortgage calculator.

Miscellaneous

2016: Updated the JHU thesis template available via GitHub and Overleaf.

Computer skills

all-purpose: R *Ranked 211/8386 (top 2.5%) in the US and 759/71400 (top 1%) worldwide by GitHub Awards as of March 14, 2018. Does not take into account contributions at LieberInstitute, leekgroup and other GitHub organizations.*

statistics: Stata

scripting: bash, Perl

markup: LaTeX, markdown

OS: Linux

cluster queue: Sun Grid Engine

Teaching Experience

Instructor

- LIBD

  2016: Biostatistics and Stata instructor at a workshop for Kandahar University Faculty, organized by Johns Hopkins University. (website) 8 enrollment held at Dubai, UAE.

  2016: Invited instructor for Genomeeting 2016 ~ 40 enrollment held at INMEGEN, Mexico City, MX.

- PDCB-UNAM, Cuernavaca, MX

  2011: Invited instructor for the course *Introduction to R and Biostatistics* (website) ~ 10 enrollment.

  2010: *Analysis of High-Throughput Sequencing data with Bioconductor* for Biomedical Sciences PhD Program students (website) ~ 10 enrollment.

- CCG-UNAM, Cuernavaca, MX

  2010: *Introduction to Using Bioconductor for High-Throughput Sequencing Analysis* practice lab at the *National Bioinformatics Week* ~ 40 enrollment.

- IBT-UNAM, Cuernavaca, MX

  2010: *Introduction to R and plotting with R* course for Morett's lab ~ 10 enrollment.

  2010: Organized and gave a lecture for the course on *Statistical Methods and Analysis of Genomic Data* (website) ~ 20 enrollment.

  2009: Organized the course *Introduction to Bioinformatics* for Morett's lab and served as instructor for the *Introduction to R and plotting with R* module (website) ~ 10 enrollment.

- LCG-UNAM, Cuernavaca, MX

  2009: *Seminar III: R/Bioconductor*. In-depth Bioconductor course (website) ~ 30 enrollment.

Guest lecturer

- JHBSPH, Baltimore, US

  2015: *Introduction to R for Public Health Researchers: Reproducible research module* ~ 20 enrollment.

- LCG-UNAM, Cuernavaca, MX

  2012: *Introduction to R and Biostatistics* lecture for *Seminar 1: Introduction to Bioinformatics* course ~ 30 enrollment.

  2011: *Introduction to R and Biostatistics* lecture for *Seminar 1: Introduction to Bioinformatics* course ~ 30 enrollment.

Lead teaching assistant

- JHBSPH, Baltimore, US
  2015–2016: *Statistical Methods in Public Health II* ~ 550 enrollment.
  2014–2015: *Statistical Methods in Public Health I* and *II* ~ 550 enrollment.

Teaching assistant

- JHBSPH, Baltimore, US
  2014–2016: *MPH capstone project*: 30 min one-on-one consulting sessions (biostatistics, Stata coding) ~ 500 enrollment. Develop and maintain the MPHcapstoneTA shiny application.
  2015–2016: *Statistical Methods in Public Health I* ~ 550 enrollment.
  2015: *Introduction to R for Public Health Researchers* ~ 20 enrollment.
  2013–2014: *Statistical Methods in Public Health I* and *II* ~ 550 enrollment.
  2012–2013: *Statistical Methods in Public Health I, II, III,* and *IV* ~ 550 enrollment.
- LCG-UNAM, Cuernavaca, MX
  2009: *Principles of Statistics*. Basic R (website) ~ 30 enrollment.
  2008: *Bioinformatics and Statistics I*. R and Bioconductor overview (website) ~ 40 enrollment.

Mentoring

2017-2018: MPH practicum and MPH capstone advisor for Amy Peterson.

2017: MPH practicum and MPH capstone advisor for Ashkaun Razmara.

2015: Mentored Alquicira-Hernández J, LCG-UNAM student visiting Jeff Leek's group.

2009–2011: Advised and trained 13 LCG-UNAM students and alumni while working at *Winter Genomics*: Riveros-McKay F, Vargas-Chávez C, Dulanto-Acevedo V, Romero-Martínez S, Samaniego-Castruita J, Zepeda-Mendoza L, Vargas-Velázquez A, Noé-González M, Soto Jiménez LM, López Moyado I, Medina-Abarca H., Izquierdo-Rangel E, and Berrocal-Quezada NA.

2009: Trained 3 LCG-UNAM students to take over the R/Bioconductor course: Reyes-Quiroz A, Moreno-Mayar V, and Reyes-López J.

Other

2018: Co-founder of CDSBMexico which is a community of R and Bioconductor developers in Latin America.

2018: Co-founder of the LIBD rstats club.

2016: Student representative for the Centennial celebration of the Department of Biostatistics.

2012–2016: Organized *Cultural Mixer* events for the Department of Biostatistics with Amanda Mejia for raising cultural awareness.

2012–2014: Organized the *Genomics for Students* group (website)

2009–2011: Organized a Genomics Journal Club at IBT-UNAM.

2008–2009: Elected class representative for the LCG Academic Committee.

2008–2009: Class representative for Administration Unit for Technology Information committee.

2008: Helped start the National Node of Bioinformatics (Mexico) online forum.

Languages

Native: Spanish

Bilingual: English

Basic: French