Leonardo Collado-Torres

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

📞 +1 (410) 955 1000 • ✉ leo.collado@libd.org
🌐 http://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

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

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

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

Experience

Research

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

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**

Scientific executive   *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

Peer-reviewed..

1. 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.
2. **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.
3. 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.

4. **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.
5. 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.
6. 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.
7. 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.

Pre-prints

1. 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 B, Maher BJ, Deep-Soboslay A, The BrainSeq Consortium, Cross A, Brandon NJ, Leek JT, Hyde TM, Kleinman JE, Weinberger DR. Developmental and genetic regulation of the human cortex 1 transcriptome in schizophrenia. *bioRxiv* 124321 (2017). doi: 10.1101/124321.
2. **Collado-Torres L***, Nellore A*, Kammers K, Ellis SE, Taub MA, Hansen KD, Jaffe AE, Langmead B, Leek JT. recount: A large-scale resource of analysis-ready RNA-seq expression data. *bioRxiv* 068478 (2016). doi: 10.1101/068478.

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
SlideShare: lcolladotor

LinkedIn: lcollado

Epernicus: lc40

Professional service

Develop and maintain open-source biostatistical software.

Peer review

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

2015-2016: American Statistical Association

2014-2016: ENAR student member

2014: American Public Health Association

Presentations

Talks at conferences

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

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

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.

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.

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.

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.

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.

Posters

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.

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.

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.

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.

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.

Other talks

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

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

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

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

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

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

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

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

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

2012: Introducing Git while making your academic webpage, *Biostatistics Computing*

Club, JHBSPH, Baltimore – US.

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

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

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

Courses and Meetings Attendance

2016: *ENAR*, Austin – US.

2015: *ENAR*, Miami – US.

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

2014: *BioC2014*, Harvard, 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: *BioC2011*, FHCRC, Seattle – US.

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

2010: *BioC2010*, FHCRC, Seattle – US.

2009: *BioC2009*, FHCRC, Seattle – US.

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

2008: *BioC2008*, 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 – 1603 downloads.

2014: derfinder: Annotation-agnostic differential expression analysis of RNA-seq data at base-pair resolution via the DER Finder approach – 8475 downloads.

2014: derfinderPlot: plotting functions for derfinder results – 5858 downloads.

2014: regionReport: Generate HTML or PDF reports for a set of genomic regions or DESeq2/edgeR results – 5934 downloads.

2014: derfinderHelper: helper functions for derfinder package – 7438 downloads.

2014: derfinderData: data for derfinder examples – 1371 downloads.

Bioconductor – contributor role..

2015: bumphunter

2014: ballgown

Other R packages..

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 at GitHub and Overleaf

Computer skills

all-purpose: R *Ranked 163/6361 in the US and 483/57692 worldwide by GitHub Awards as of December 19, 2016. Does not take into account contributions at LieberInstitute and leekgroup 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.

○ 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

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

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 online forum.

Languages

Native: Spanish

Bilingual: English

Basic: French