Leonardo Collado-Torres

Education

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

2005-2009 **Bachelor in Genomic Sciences (LCG)**, *National Autonomous University of Mexico (UNAM)*, Cuernavaca, MX.

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

PhD thesis

title *Developing single-base RNA-seq methods and software to understand neuropsychiatric disorders and more*

advisors Jeffrey T. Leek and Andrew E. Jaffe

description The goal is to develop statistical methods and software that can enable researchers to differentiate as much as possible the sources of variation observed in RNA-seq while minimizing the dependance on known annotation. Thus allowing researchers to correct for technological variation and study the biological variation driving their phenotype of interest. Then apply these methods to further our understanding of neuropsychiatric disorders using the Lieber Institute for Brain Development data collection.

Honors and awards

2011 Awarded CONACyT Mexico scholarship for PhD studies outside Mexico

2009 Honorable Mention for bachelor studies at LCG-UNAM

2005 Best high school average: awarded ITESM system 90% scholarship for college studies, declined to join LCG-UNAM

Experience

Vocational

2009-2011 **Scientific executive**, *Winter Genomics*, Cuernavaca, MX.

Responsible for recruiting and hiring new personnel, overseeing and supervising nine 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,
  - de novo assembly of four *Escherichia coli* strains and lead to Aguilar et al, PMID 22884033;
- Designed training material for new employees.

Research

2009–2011 **Bioinformatician**, *Enrique Morett lab*, IBT-UNAM, Cuernavaca, MX.

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

2007–2009 **Undergraduate research assistant**, *Guillermo Dávila lab*, CCG-UNAM, Cuernavaca, MX.

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

2007 **Undergraduate research assistant**, *Roberto Kolter lab*, Harvard, Boston, US.

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*.

Publications

Peer-reviewed

1. 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.* (2014). doi:10.1038/nn.3898.

2. 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.

3. 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.

Pre-prints

1. **Collado-Torres L**, Frazee AC, Love MI, Irizarry RA, Jaffe AE, Leek JT. *derfinder: Software for annotation-agnostic RNA-seq differential expression analysis*. bioRxiv 015370 (2015). doi:10.1101/015370.

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.

Google scholar

`https://scholar.google.com/citations?user=h57-MykAAAAJ`

Referee for

2013-2014 Biostatistics

Presentations

Talks

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

2014 **Collado-Torres L**, Frazee AC, Love MI, Irizarry RA, Jaffe AE, Leek JT. Fast annotation-agnostic differential expression analysis, *ENAR*, Baltimore – US.

2013 **Collado-Torres L**, Jaffe AE, Leek JT. Fast annotation-agnostic differential expression analysis, *Genomics and Bioinformatics Symposium*, JHSPH, 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.

Teaching Experience

Instructor

○ PDCB-UNAM, Cuernavaca, MX

2011 Invited instructor for the course *Introduction to R and Biostatistics* `http://www.biostat.jhsph.edu/~lcollado/PDCB-Biostats.html` ~ 10 enrollment.

2010 *Analysis of High-Throughput Sequencing data with Bioconductor* for Biomedical Sciences PhD Program students `http://www.biostat.jhsph.edu/~lcollado/PDCB-HTS.html` ~ 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* `http://www.biostat.jhsph.edu/~lcollado/MEyAdDG.html` ~ 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 `http://www.biostat.jhsph.edu/~lcollado/mIntroR.html` ~ 10 enrollment.

○ LCG-UNAM, Cuernavaca, MX

2009 *Seminar III: R/Bioconductor*. In-depth Bioconductor course `http://www.lcg.unam.mx/~lcollado/B/` ~ 30 enrollment.

Lead teaching assistant

○ JHSPH, Baltimore, US

2014–2015 *Statistical Methods in Public Health I* and *II* ~ 550 enrollment.

Teaching assistant

○ JHSPH, Baltimore, US

2015 *MPH capstone project*: 30 min one-on-one consulting sessions (biostatistics, Stata coding) ~ 500 enrollment.

2014 *MPH capstone project*: 30 min one-on-one consulting sessions (biostatistics, Stata coding) ~ 500 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 `http://www.lcg.unam.mx/~lcollado/E/` ~ 30 enrollment.

2008 *Bioinformatics and Statistics I*. R and Bioconductor overview `http://www.lcg.unam.mx/~lcollado/R/` ~ 40 enrollment.

Courses and Meetings Attendance

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

2014 *BioC2014*, Harvard, Boston – US.

2014 *ENAR*, Baltimore – US.

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

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

2015 bumphunter: contributor role.

2014 derfinder: Annotation-agnostic differential expression analysis of RNA-seq data at base-pair resolution – main author.

2014 derfinderPlot: plotting functions for derfinder results – main author.

2014 regionReport: generate HTML reports for exploring a set of regions – main author.

2014 derfinderHelper: helper functions for derfinder package – main author.

2014 derfinderData: data for derfinder examples – main author.

2014 ballgown: contributor role.

Other R packages

2014   dots: simplify function calls `https://github.com/lcolladotor/dots`.

2013   fitbitR: visualize your FitBit data `https://github.com/lcolladotor/fitbitR`.

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

shiny applications

2014–2015   MPH capstone TA office hours sign up `http://glimmer.rstudio.com/lcolladotor/MPHcapstoneTA/`

2014   Simple mortgage calculator `https://lcolladotor.shinyapps.io/mortgage2/`

Computer skills

all-purpose	**R** *Ranked 93/4371 in the US and 227/39364 worldwide by* `http://github-awards.com/users/lcolladotor` *as of March 08, 2015*
statistics	**Stata**
scripting	**bash, Perl**
markup	**LaTeX, markdown**
OS	**Linux**

Languages

Native	Spanish
Bilingual	English
Basic	French

Other

2012–2014   Organized the *Genomics for Students* group `http://www.biostat.jhsph.edu/~lcollado/GenomicsForStudents.html`

2012–2014   Organized *Cultural Mixer* events for the Department of Biostatistics (JHSPH) with Amanda Mejia.

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 `http://foros.nnb.unam.mx/`.