Harold Pimentel

Postdoctoral researcher
HHMI Hanna H. Gray fellow

Contact Information

Cell:	(111) 222-3333
E-mail:	haroldpimentel@gmail.com
Web:	http://pimentel.github.io
GitHub:	http://github.com/pimentel
Blog:	http://haroldpimentel.wordpress.com

Education

University of California, Berkeley, Berkeley, California, USA

Ph.D., Computer Science **August 2010 - August 2016**

- Advisor: Prof. Lior Pachter
- Thesis: ***Fast and accurate quantification and differential analysis of transcriptomes***.
- Thesis committee: Haiyan Huang, Lior Pachter, Nir Yosef
- Qualifying exam committee (May 2014): Doris Bachtrog, Haiyan Huang, Lior Pachter, Nir Yosef
- Graduated with Designated Emphasis in Computational and Genomic Biology
- Specialization: Computational biology, (bio)statistics, machine learning
- Supported by NSF Graduate Research Fellowship Program (2011-2014)
- Supported by GAANN fellowship, and Academic Excellence Award (2010-2011)

M.A., Statistics **December 2013**

- Advisor: Prof. Haiyan Huang
- Specialization: Applied statistics
- Thesis: ***Biclustering as an application of sparse canonical correlation analysis with RNA-Seq applications***.
- Thesis committee: Peter J. Bickel, Haiyan Huang, Lior Pachter

California State University, Fullerton, Fullerton, California, USA

B.S., Computer Science **August 2005 - May 2010**

- Specialization: Scientific computing
- Minor: Mathematics
- Graduated with summa cum laude

Employment

Stanford University, Stanford, California USA

Postdoctoral Researcher **August 2016 - Present**

Professor Jonathan Pritchard Lab. Interested in RNA expression, splicing, regulation, and how genetic variation affects those. Focused on statistical and algorithmic method development in: intron retention QTL mapping, modeling of targeted perturbations, network-guided trans-eQTL mapping. Additionally, ongoing collaboration with Prof. Hiten Madhani (UCSF) on developing statistical techniques for modeling conservation of methylation in highly repetitive regions (e.g. centromeres) in species *C. neoformans*.

University of California, Berkeley, Berkeley, California USA

Research Assistant **October 2010 - July 2016**

Professor Lior Pachter Lab. Generally focused on method development using statistics and machine learning for mRNA-Seq technology. Work included: identifying incorrect gene annotations, differential expression coupled with transcriptome abundance estimation, transcriptome mapping, visualizing fragment alignments, and analysis with biologist collaborators. Also the system administrator for the group.

Research Assistant **June 2013 - July 2016**

Professor Haiyan Huang Lab. Developing methods for biclustering of linear expression patterns by sparse canonical correlation analysis. Our method generalizes many existing biclustering methods.

10X Genomics, Pleasanton, California USA

Computational biology intern   **May 2014 - August 2014**
Using statistical and machine learning techniques (MART, GLM) to analyze biological sequencing data from a novel sequencing platform. Also helped develop analysis pipelines in Python and R.

California State University, Fullerton (CSUF), Fullerton, California USA

Research Assistant   **December 2007 - May 2010**
Professor Spiros Courellis Lab. Shake&wARP electron density map refinement research project. Deployed a distributed and parallel computing environment used for Phase Bias Removal in Macromolecular Crystallography. Developed new techniques for alternative electron density maps, including new methods to cluster maps based on similarity and returning a new averaged map based on the most similar maps.

Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Summer Research Intern   **June 2009 - August 2009**
Professor David Gifford Lab. Developed and implemented a technique for determining the affinity of a transcription factor to a repetitive element family in ChIP-seq reads. Application was used to test the affinity of Retinoic Acid Receptor to ALU repetitive elements.

Thales Raytheon Systems, Fullerton, California USA

Information Solutions Intern   **May 2007 - May 2008**
Provided reliable, innovative, effective, and efficient computing infrastructure to employees. Tasks included support for PGP Encryption, server monitoring, system administration, etc.

Publications

Full list of publications at Google Scholar.

≫ Denotes highlighted publication/presentation.

H. Pimentel, Y. Li, L. Pachter, J. K. Pritchard: "Fast intron retention QTL analysis", ***In preparation. Pre-print expected October 2019.***

≫ S. Catania, *P. A. Dumesic, *H. Pimentel, A. Nasif, C. I. Stoddard, J. E. Burke, J. K. Diedrich, S. Cook, T. Shea, E. Geinger, R. Lintner, J. R. Yates III, P. Hajkova, G. J. Narlikar, C. A. Cuomo, J. K. Pritchard, H. D. Madhani: "Evolutionary persistence of DNA methylation for millions of years after ancient loss of a de novo methyltransferase", ***bioRxiv (2019) In review.*** doi:10.1101/149385 *contributed equally

H. Pimentel, Z. Hu, H. Huang: "Biclustering by by sparse canonical correlation analysis", ***Quantitative Biology (2018)***. doi:10.1007/s40484-017-0127-0

L. Yi, H. Pimentel, N. Bray, L. Pachter: "Gene-level differential analysis at transcript-level resolution", ***Genome Biology (2018)***. doi:10.1186/s13059-018-1419-z

A. C. Smart, C. A. Margolis, H. Pimentel, M. X. He, D. Miao, D. Adeegbe, T. Fugmann, K. Wong, E. M. Van Allen: "Intron retention is a source of neoepitopes in cancer", ***Nature Biotechnology (2018)***. doi:10.1038/nbt.4239

≫ H. Pimentel, N. Bray, S. Puente, P. Melsted, L. Pachter: "Differential analysis of RNA-Seq incorporating quantification uncertainty", ***Nature Methods 14, 7, 687 (2017)***. doi:10.1038/nmeth.4324

L. Schaeffer, H. Pimentel, N. Bray, P. Melsted, L. Pachter: "Pseudoalignment for metagenomic read assignment", ***Bioinformatics (2017)***. doi:10.1093/bioinformatics/btx106

L. Yi, H. Pimentel, L. Pachter: "Zika infection of neural progenitor cells perturbs transcription in neurodevelopmental pathways", ***PLOS ONE (2017)***, doi:10.1371/journal.pone.0175744

H. Pimentel, P. Sturmfels, N. Bray, P. Melsted, L. Pachter. "The Lair: a resource for exploratory analysis of published RNA-Seq data", ***BMC Bioinformatics (2016)***, doi:10.1186/s12859-016-1357-2

≫ N. Bray, H. Pimentel, P. Melsted, L. Pachter: "Near-optimal probabilistic RNA-Seq quantification", ***Nature Biotechnology (2016)*** doi:10.1038/nbt.3519

≫ H. Pimentel, M. Parra, S. Gee, N. Mohandas, L. Pachter, J. G. Conboy: "A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis", ***Nucleic Acids Research (2015)***, doi: 10.1093/nar/gkv1168

H. Pimentel, J. G. Conboy, L. Pachter: "Keep Me Around: Intron Retention Detection and Analysis", *aXiv preprint* (2015), http://arxiv.org/abs/1510.00696

H. Pimentel, M. Parra, S. Gee, D. Ghanem, X. An, J. Li, N. Mohandas, L. Pachter, J. G. Conboy: "A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis", ***Nucleic Acids Research (2014)***, doi:10.1093/nar/gkt1388

D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions", ***Genome Biology 14:R36 (2013)***

C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter: "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks" ***Nature Protocols 7, 562578 (2012)***

A. Roberts, H. Pimentel, C. Trapnell, L. Pachter: "Identification of novel transcripts in annotated genomes using RNA-Seq", ***Bioinformatics (2011) 27 (17): 2325-2329.***

PRESENTATIONS

≫ H. Pimentel, Y. Li, L. Pachter, J. K. Pritchard: "Fast intron retention QTL analysis", ***Probabilistic Modeling in Genomics***, Cold spring Harbor Laboratory, USA (November, 2018; Oral presentation)

H. Pimentel, Y. Li, L. Pachter, J. K. Pritchard: "Fast intron retention QTL analysis", ***HHMI symposium***, HHMI Janelia Research Campus, USA (November, 2018; Poster presentation)

≫ *H. Pimentel, *N. Bray, P. Melsted, L. Pachter: "Ultrafast RNA-seq Analysis with Kallisto and Sleuth", ***Functional Genomics Seminar***, UC Davis, USA (February, 2016; Invited by Sharon Aviran). *denotes joint presentation

H. Pimentel, N. Bray, P. Melsted, L. Schaeffer, L. Pachter: "Transcript level abundance estimation and differential expression with kallisto and sleuth", ***RNA-Seq symposium***, Janssen Pharmaceuticals, WebEx (November, 2015)

H. Pimentel, N. Bray, P. Melsted, L. Pachter: "Transcript level differential analysis with sleuth", ***Genome Informatics***, Cold Spring Harbor Laboratory, USA (October, 2015; Poster Presentation).

≫ *H. Pimentel, *N. Bray, P. Melsted, L. Pachter: "Ultrafast RNA-seq Analysis with Kallisto and Sleuth", ***Systems Biology Seminar***, UCSF, USA (July, 2015; Invited by Raj Bhatnagar). *denotes joint presentation

H. Pimentel, N. Bray, P. Melsted, L. Pachter: "Transcript-level quantification and differential expression with measurement uncertainty using RNA-Seq", ***Ramsingh Lab***, University of Southern California, USA (June, 2015; Ramsingh Group Presentation).

N. Bray, H. Pimentel, P. Melsted, L. Pachter: "Ultrafast accurate RNA-Seq analysis", ***Biology of Genomes***, Cold Spring Harbor Laboratory, USA (June, 2015; Poster Presentation).

≫ <u>H. Pimentel</u>, M. Parra, S. Gee, N. Mohandas, L. Pachter, J. G. Conboy: "An Erythroid-Specific Intron Retention Program Regulates Expression of Selected Genes during Terminal Erythropoiesis", ***American Society of Hematology 2014***, San Francisco, California, USA (December, 2014; Oral Presentation).

<u>H. Pimentel</u>, H. Huang: "Biclustering as an application of sparse canonical correlation analysis", ***ISMB 2014***, Boston, Massachusetts, USA (July, 2014; Poster Presentation). Travel fellowship provided by FASEB MARC.

<u>H. Pimentel</u>, A. Roberts, C. Trapnell, L. Pachter: "Visualizing RNA-Seq fragment alignments", ***ISMB 2011***, Vienna, Austria (July, 2011; Poster Presentation). Travel fellowship provided by FASEB MARC.

<u>H. Pimentel</u>, S. H. Courellis, K. A. Kantardjieff, B. Rupp: "Extending Shake&wARP to Allow the Selection of the Refined Protein Electron Density Map from a Set of Sub-optimal Solutions." ***CSU Program for Education and Research in Biotechnology 2010***, Los Angeles, CA. (January 2010; Poster Presentation)

<u>H. Pimentel</u>, S. H. Courellis, K. A. Kantardjieff, B. Rupp: "Performance Evaluation of Shake&wARP, a Cluster Enabled Application for Macromolecular Crystallography." ***International Conference of Computing in Engineering, Science, and Information 2009***, Fullerton, CA. (April 2009; Oral Presentation and published in the IEEE Proceedings.)

<u>H. Pimentel</u>, S. H. Courellis, K. A. Kantardjieff, B. Rupp: "Hosting Multiple Simultaneous Applications over an Easy to Administrate, Scalable, Extensible Platform Integrating Diverse Computational Resources." ***Computer Science and Information Engineering 2009***, Los Angeles, CA. (April 2009; Poster Presentation and published in the IEEE Proceedings vol.3, no., pp.67-71)

<u>H. Pimentel</u>, S. H. Courellis, K. A. Kantardjieff, B. Rupp: "A performance evaluation of Shake&wARP, an application used in macromolecular crystallography for electron density map refinement." ***SACNAS National Conference 2009***, Dallas, TX. (October 2009; Poster Presentation)

<u>H. Pimentel</u>, S. Mahony, D. K. Gifford: "Determining a Transcription Factor's Affinity to Repetitive DNA." ***MIT Summer Research Program Poster Session 2009***, MIT, Cambridge, MA. (August 2009; Poster Presentation)

<u>H. Pimentel</u>, S. H. Courellis, K. A. Kantardjieff, B. Rupp: "A high throughput implementation of Shake&wARP, a Software Application used for Phase Bias Removal in Macromolecular Crystallography." ***Natural Science and Mathematics Career Day 2009***, CSUF, Fullerton, CA. (April 2009; Poster Presentation)

<u>H. Pimentel</u>, S. H. Courellis, K. A. Kantardjieff, B. Rupp: "A high throughput implementation of Shake&wARP, a Software Application used for Phase Bias Removal in Macromolecular Crystallography." ***CSU Program for Education and Research in Biotechnology 2009***, Los Angeles, CA. (January 2009; Poster Presentation)

SELECTED HONORS AND AWARDS

HHMI Hanna H. Gray Fellowship **September 2017 - September 2023**
Transitional award (Postdoc to PI; $1.4M over 8 years)

NIH Supplement to Promote Diversity in Health-Related Research **November 2016 - September 2017**

FASEB MARC Travel Award for ISMB 2014 **August July 2014**

National Science Foundation Graduate Research Fellowship Program (NSF GRFP) **August 2011 - May 2015**

FASEB MARC Travel Award for ISMB 2011 **August July 2011**

Graduate Assistance in Areas of National Need (GAANN) Fellow **August 2010 - May 2011**

NIH Minority Access to Research Careers (MARC) Scholar **June 2008 - May 2010**

Boeing Dean's Scholarship **April 2009 - 2010**

"Outstanding poster presentation in Computer Science" at SACNAS Conference **October 2009**

Southern California Auto Scholarship **August 2008 - 2009**

OPEN SOURCE SOFTWARE

Academic:
- sleuth (2015 - present) - Transcript level differential expression with uncertainty estimation (R)
  https://github.com/pachterlab/sleuth
- kallisto (2015) - Near optimal RNA-Seq quantification (C++11)
  https://github.com/pachterlab/kallisto
- Keep Me Around (2015) - Intron Retention Detection and Analysis (R)
  https://github.com/pachterlab/kma
- Fast SCCA (2014) - Implementation of SCCA using the NIPALS algorithm (C++, R)
  https://github.com/pimentel/fscca
- TopHat (2011) - Modules for transcriptome mapping since version 1.4.0 (C++)
  http://ccb.jhu.edu/software/tophat

Non-academic:
- vim-slimeblocks (2018) - Send R/python code to tmux sockets.
  https://github.com/pimentel/vim-slimeblocks
- atom-r-exec (2016) - Send R code to various R consoles from the Atom editor (CoffeeScript)
  https://github.com/pimentel/atom-r-exec
- Other code can be found at http://github.com/pimentel

COMPUTER SKILLS
- Languages:
  - Proficient: R, C++, Python
  - Competent: JavaScript, SQL, Bash
- Several other "retired" languages which I haven't used in a while, but was once fluent in: csh, Java, Perl, MATLAB, Fortran
- Unit testing
- System administration in Linux/UNIX
- Tools: Git, Vim, CMake, Make, Snakemake, Subversion, OpenPBS/Torque

Member of Stanford Latinx Postdoc Association (June 2017 - Present)

Member of Berkeley Science Network (January 2013 - July 2016)

President for Association of Computing Machinery, CSUF Chapter (August 2009 - May 2010)

President of the Omicron (CSUF) chapter for Upsilon Pi Epsilon, international computer science honor society (January 2008 - May 2009)

Vice President for Association of Computing Machinery, CSUF Chapter (January 2008 - May 2009)

Member of Orange County Linux Users Group as well as UNIX Users Association of Southern California (June 2007 - May 2010)

TEACHING

GSI for "Introduction to Statistical Computing" (STAT 243). Main course taught by Chris Paciorek. (Fall 2015)

GSI for two sections of "Structure and Interpretation of Computer Programs" (CS 61A). Main course taught by Professor John DeNero. (Spring 2015)

MENTORING

Alex Tseng (Undergraduate; August 2015 - August 2016) has been working on improving effective length estimation in RNA-Seq transcript quantification.

Pascal Sturmfels (Undergraduate; August 2015 - May 2016) has been working on improving the memory usage of sleuth so that sleuth objects can be shared easily for easier reproducibility and accessibility.

Diego Alcantar (Undergraduate; September 2015 - May 2016) has been working on integrating pathway analysis into sleuth.

Daniel Li (Undergraduate; August 2015 - December 2015) worked on implementing visualizations for principal component analysis loadings and bias parameters in sleuth.

WORKSHOPS

"Differential expression analysis with Sleuth." *Kallisto-Sleuth workshop at UC Berkeley*. Hosted by Laboratory of Mathematics and Computational Biology (2016). Slides at https://pachterlab.github.io/kallisto-sleuth-workshop-2016/

"Algorithms for RNA-Seq" *iPlant workshop at UC Davis*. Hosted by iPlant Collaborative (2015). Video at https://youtu.be/96yBPM8lEt8?list=PLfFNmoa-yUIb5cYG2R1zf5rtrQQKZvKwG

"Algorithms for RNA-Seq" *RNA-Seq for the Next Generation*. Hosted by Cold Spring Harbor Laboratory (2015). Slides at http://www.rnaseqforthenextgeneration.org/

"eXpress transcript quantification" **Seq I Workshop*. Hosted by Laboratory of Mathematics and Computational biology (2012). http://lmcb.wikispaces.com/eXpress+Walkthrough

"Sequence alignment" *RNA-Seq workshop*. Hosted by Computational Genomics Resource Laboratory (2011). http://cgrlucb.wikispaces.com/Sequence+alignment

OUTREACH

Presentations to biology and statistics classrooms at La Mirada High School on computational biology (Spring 2014)

Mentor to summer research intern through SUPERB program at UC Berkeley (Summer 2011)

Mathematics, Engineering, Science Achievement (MESA) Robotics Workshop (April, May 2009)

HOBBIES

Sketching, fountain pens and caligrpahy, road and mountain cycling, bicycle mechanics.