

Harold Pimentel

CONTACT INFORMATION

Cell: (999) 999-9999
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. candidate, Computer Science

Fall 2010 - Present

- Advisor: Prof. Lior Pachter
- Qualifying exam committee (May 2014): Doris Bachtrog, Haiyan Huang, Lior Pachter, Nir Yosef
- 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

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

PROFESSIONAL EXPERIENCE

10X Genomics, Pleasanton, California USA

Computational biology intern

May 2014 - August 2014

Using statistics 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.

University of California, Berkeley, Berkeley, California USA

Research Assistant

October 2010 - Present

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

Research Assistant

June 2013 - Present

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

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.

HONORS AND AWARDS

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

Outstanding Sophomore, Junior, and Senior at CSUF - Highest class GPA

**May 2009
2008
2007**

Computer Science Dean's List at CSUF

Fall 2005 - Spring 2010

PUBLICATIONS

N. Bray, H. Pimentel, P. Melsted, L. Pachter: "Near-optimal RNA-Seq quantification", *arXiv preprint* (2015), <http://arxiv.org/abs/1505.02710>

H. Pimentel, H. Huang: "Generalized linear biclustering by sparse canonical correlation analysis", *Manuscript in preparation*.

H. Pimentel, J. G. Conboy, L. Pachter: "Keep Me Around: Intron Retention Detection and Analysis", *Manuscript under review*.

H. Pimentel, 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). *Accompanying manuscript in preparation*.

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, 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).

H. Pimentel, 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.

H. Pimentel, 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.

H. Pimentel, 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)

H. Pimentel, 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.)

H. Pimentel, 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)

H. Pimentel, 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)

H. Pimentel, 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)

H. Pimentel, 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)

H. Pimentel, 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)

PROFESSIONAL AFFILIATIONS	<p>Member of Berkeley Science Network (January 2013 - Present)</p> <p>President for Association of Computing Machinery, CSUF Chapter (August 2009 - May 2010)</p> <p>President of the Omicron (CSUF) chapter for Upsilon Pi Epsilon, international computer science honor society (January 2008 - May 2009)</p> <p>Vice President for Association of Computing Machinery, CSUF Chapter (January 2008 - May 2009)</p> <p>Member of Orange County Linux Users Group as well as UNIX Users Association of Southern California (June 2007 - May 2010)</p>
OPEN SOURCE SOFTWARE	<ul style="list-style-type: none"> • kallisto (2015) - Near optimal RNA-Seq quantification (C++11) http://github.com/pachterlab/kallisto • Keep Me Around (2015) - Intron Retention Detection and Analysis (R) http://github.com/pachterlab/kma • Fast SCCA (2014) - Implementation of SCCA using the NIPALS algorithm (C++, R) http://github.com/pimentel/fscqa • TopHat (2011) - Modules for transcriptome mapping since version 1.4.0 (C++) http://ccb.jhu.edu/software/tophat • Other code can be found at http://github.com/pimentel
COMPUTER SKILLS	<ul style="list-style-type: none"> • Languages: <ul style="list-style-type: none"> • Proficient: R, Python • Competent: C++, SQL • Advanced beginner: 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
TEACHING	<p>GSI for “Introduction to Statistical Computing” (STAT 243). Main course taught by Chris Paciorek. (Fall 2015)</p> <p>GSI for two sections of “Structure and Interpretation of Computer Programs” (CS 61A). Main course taught by Professor John DeNero. (Spring 2015)</p>
WORKSHOPS	<p>“Algorithms for RNA-Seq” <i>RNA-Seq for the Next Generation</i>. Hosted by Cold Spring Harbor Laboratory (2015). Slides posted shortly to http://www.rnaseqforthenextgeneration.org/</p> <p>“eXpress transcript quantification” <i>*Seq I Workshop</i>. Hosted by Laboratory of Mathematics and Computational biology (2012). http://lmcw.wikispaces.com/eXpress+Walkthrough</p> <p>“Sequence alignment” <i>RNA-Seq workshop</i>. Hosted by Computational Genomics Resource Laboratory (2011). http://cgrlucb.wikispaces.com/Sequence+alignment</p>
OUTREACH	<p>Presentations to biology and statistics classrooms at La Mirada High School on computational biology (Spring 2014)</p> <p>Mentor to summer research intern through SUPERB program at UC Berkeley (Summer 2011)</p> <p>Mathematics, Engineering, Science Achievement (MESA) Robotics Workshop (April, May 2009)</p>
HOBBIES	<p>Road and mountain cycling, bicycle mechanics, and hiking</p>