

# Harold Pimentel

---

## CONTACT INFORMATION

*Cell:* (999) 999-9999  
*E-mail:* haroldpimentel@gmail.com  
*WWW:* <https://pimentel.github.io>

## OBJECTIVE

To find an internship where I can apply my statistics, machine learning, and computer science skills. In general, I am interested in data science, especially analyzing high-dimensional data.

## EDUCATION

**University of California, Berkeley**, Berkeley, California, USA

Ph.D. candidate, Computer Science

**Fall 2010 – Fall 2015 (expected)**

- Advisor: Prof. Lior Pachter
- Specialization: Computational biology, (bio)statistics, machine learning
- Supported by NSF Graduate Research Fellowship Program (2011–2015)
- 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. Development of data analysis methods using statistics and machine learning for mRNA-Seq data. Currently working on identifying incorrect gene annotations using linear programming, differential expression coupled with transcriptome abundance estimation using shrinkage estimators, and data analysis with biologist collaborators. Past work: transcriptome mapping algorithms, 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 extending sparse canonical correlation analysis. Our method generalizes many existing biclustering methods. We are focusing on applications to large heterogeneous gene expression experiments and time series gene expression experiments.

**California State University, Fullerton**, Fullerton, California USA

*Research Assistant*

**December 2007 – May 2010**

Professor Spiros Courellis Lab. Deployed a distributed and parallel computing environment used for Phase Bias Removal in Macromolecular Crystallography. Developed 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.

OPEN SOURCE  
SOFTWARE

- 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>
- Miscellaneous code can be found at <http://github.com/pimentel>

COMPUTATIONAL  
SKILLS

- Languages:
  - Proficient: R, Python
  - Competent: C++, SQL
  - Advanced beginner: Bash
  - Experience with but currently rarely use: Java, Perl, MATLAB, Fortran, csh
- Unit testing
- System administration in Linux/UNIX
- Tools: Git, CMake, Vim, OpenPBS/Torque

SELECTED  
PRESENTATIONS  
AND PUBLICATIONS  
(SINCE 2010)

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).

H. Pimentel, H. Huang: “Biclustering as an application of sparse canonical correlation analysis”, *ISMB 2014*, Boston, MA, USA (July, 2014; Poster Presentation). Travel fellowship provided by FASEB MARC.

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 (Journal Publication)

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)* (Journal Publication)

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

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.

HONORS AND  
AWARDS  
(SINCE 2010)

FASEB MARC Travel Award for ISMB 2014

July 2014

National Science Foundation Graduate Research Fellowship  
Program (NSF GRFP)

August 2011 – May 2015

FASEB MARC Travel Award for ISMB 2011

July 2011

Graduate Assistance in Areas of National Need (GAANN) Fellow

August 2010 – May 2011

OUTREACH  
(SINCE 2010)

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)

HOBBIES

Road and mountain cycling, bicycle mechanics, and hiking