Bo Li, Ph.D.

Address
Broad Institute of MIT and Harvard
415 Main Street
Cambridge, MA 02142, USA

Contact
Phone: (608) 609-3735
Email: libo@broadinstitute.org
Homepage: http://bli25broad.github.io

Academic Employment

Jul. 2017 — present **Broad Institute of MIT and Harvard**
Postdoctoral Associate
Supervisor: Aviv Regev
Projects leading/co-leading:
1) scCloud, a cloud-based single-cell data analysis pipeline to scale up the analysis to millions of cells;
2) Immune cell atlas, profiled the gene expressions of 1.7 million single immune cells from bone marrow, cord blood, and peripheral blood;
3) Blood cell atlas, collaborated with Chloe, aiming at mapping all cell types and states in peripheral blood;
4) nuclei-hashing, a protocol to multiplex single nuclei RNA-Seq samples using antibody-oligonucleotide conjugates.

Aug. 2013 — Jun. 2017 **University of California at Berkeley**
Postdoctoral Researcher
Supervisor: Lior Pachter

Education

Sept. 2008 — Jul. 2013 **University of Wisconsin–Madison**
Ph.D. in Computer Science
Thesis: Computational analysis of RNA-Seq data in the absence of a known genome
Advisor: Colin Dewey

Sept. 2004 — Jul. 2008 **Shanghai Jiao Tong University**
B.E. in Computer Science and Engineering (ACM Honor Class)

Representative Publications

1. Gaublomme JT*, **Li B***, McCabe C, Knecht A, Drokhlyansky E, Van Wittenberghe N, Waldman J, Dionne D, Nguyen L, De Jager P, Yeung B, Zhao X, Habib N, Rozenblatt-Rosen O and Regev A. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. *bioRxiv*, 10.1101/476036, 2018.

This paper describes a novel experimental protocol to pool single nuclei from multiple samples and a novel algorithm, demuxEM, to demultiplex the pooled samples.

2. **Li B**, Tambe A, Aviran S and Pachter L. PROBer provides a general toolkit for analyzing sequencing-based toeprinting assays. *Cell Systems*, 4(5):568–574, 2017.

This paper describes PROBer – the first unified probabilistic framework for the analysis of a diverse set of sequencing-based 'toeprinting' assays. These assays are used to probe RNA secondary structure (DMS/SHAPE-Seq), detect epitranscriptomic mark (Pseudo-Seq), or identify RNA-protein interaction (iCLIP/eCLIP), which are important to understanding post-transcriptional gene regulation from all aspects.

Representative Publications (continued)

3. **Li B***, Fillmore N*, Bai Y, Collins M, Thomson JA, Stewart R and Dewey CN. Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. *Genome Biology*, 15(12):553, 2014. Highly accessed. (* Equal contribution, Citation: **139**)

This paper describes the first principled method for evaluating de novo transcriptome assemblies without ground truth.

4. **Li B** and Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12:323, 2011. Highly accessed. (Citation: **4,304**)

This paper describes the RSEM software – a widely-used RNA-Seq transcript quantification tool that is actively serving in nationwide projects such as ENCODE (The Encyclopedia of DNA Elements) and TCGA (The Cancer Genome Altas).

5. **Li B**, Ruotti V, Stewart RM, Thomson JA and Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, 26(4):493–500, 2010. (Citation: **616**)

This paper describes the RSEM algorithm.

Other Publications

1. Choudhary K, Shih NP, Deng F, Ledda M, **Li B** and S. Aviran. Metrics for rapid quality control in RNA structure probing experiments. *Bioinformatics*, 32(23):3575–3583, 2016.

2. Zeng X, **Li B**, Welch R, Rojo C, Zheng Y, Dewey CN and Keleş S. Perm-seq: Mapping protein-DNA interactions in segmental duplication and highly repetitive regions of genomes with prior-enhanced read mapping. *PLoS Computational Biology*, 11(10):e1004491, 2015.

3. Haas B, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, **Li B**, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N and Regev A. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8):1494–1512, 2013. (Citation: **2,370**)

This paper describes the widely-used Trinity de novo transcriptome assembly pipeline. RSEM contributes to the pipeline as its transcript quantification tool.

4. Chung D, Kuan PF, **Li B**, Sanalkumar R, Liang K, Bresnick EH, Dewey C and Keleş S. Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. *PLoS Computational Biology*, 7(7):e1002111, 2011.

This paper describes CSEM, one of the first ChIP-Seq multi-mapping read allocators. This work was highlighted in Nature Reviews Genetics.

Invited Talks

Assistant professor job talk, Department of Rheumatology, Allergy and Immunology, Massachusetts General Hospital, Charlestown, MA, USA. *Immunology Research in the Era of Single-Cell Genomics: Lessons Learned from 1.7 Million Single Immune Cells.* (11/2018)

Assistant professor job talk, Department of Computer Science, Duke University, Durham, NC, USA. *Taming Big Sequencing Data for RNA Biology: From Transcript Abundance Estimation to 'Epitranscriptomic' Mark Detection.* (04/2017)

Assistant professor job talk, Department of Computer Science, University of Illinois at Urbana–Champaign, Urbana, IL, USA. *Taming Big Sequencing Data for RNA Biology: From Transcript Abundance Estimation to 'Epitranscriptomic' Mark Detection.* (03/2017)

Invited Talks (continued)

Research assistant professor job talk, Toyota Technological Institute at Chicago, Chicago, IL, USA. *Taming Big Sequencing Data for RNA Biology: From Transcript Abundance Estimation to 'Epitranscriptomic' Mark Detection.* (02/2017)

The Center for RNA System Biology (CRSB) 4th Annual & Advisory Meeting with Poster Session, Berkeley, CA, USA. *PROBer: A General Toolkit for Analyzing Sequencing-based 'Toeprinting' Assays.* (10/2016)

RNA 2016: The 21st Annual Meeting of the RNA Society, Kyoto, Japan. *PROBer: A General Toolkit for Analyzing Sequencing-based 'Toeprinting' Assays.* (06/2016)

NGS Data Analysis and Informatics Conference, San Diego, CA, USA. *Quantifying RNA Information from Transcriptome-wide Chemical Probing Experiments.* (02/2016)

Core Skills in Computational Biology, Center for Computational Biology, UC Berkeley, Berkeley, CA, USA. *RNA-Seq Transcript Quantification with RSEM: A Detailed Tutorial with Common Use Cases.* (11/2015)

Teaching

University of Wisconsin–Madison
Department of Biostatistics & Medical Informatics

Fall 2008 **Teaching Assistant**, Introduction to Bioinformatics
Introducing classic computational biology algorithms for undergraduates. Responsible for office hour, homework & exam grading, and test case design for programming assignments.

Shanghai Jiao Tong University
Department of Computer Science and Engineering

Spring 2007 **Teaching Assistant**, Design and Implementation of Modern Compilers
Helping computer science majors to implement a compiler from scratch for the Tiger language. Responsible for lecturing and final project grading.

Fall 2006 **Instructor**, Programming Bootcamp: Data Structures and Algorithms
Helping computer science majors to learn data structures and algorithms by solving programming challenges. Responsible for course design, lecturing, and assignment & exam grading.

Honors and Awards

2013 Finalist, Lane Fellows in Computational Biology, Carnegie Mellon University

2010 — 2012 Morgridge Institute for Research support for senior graduate students in Computation and Informatics in Biology and Medicine

2008 Alumni Scholarship, Computer Sciences Department, University of Wisconsin–Madison

Professional Service

Reviewer: BMC Bioinformatics, Bioinformatics, and Nucleic Acids Research

Program Committee: RECOMB-Seq 2016, IJCAI 2016