

Wei Li

Department of Biostatistics and Computational Biology
Dana-Farber Cancer Institute, Harvard School of Public Health
450 Brookline Ave., Mail CLS 11007, Boston, MA 02215
wli@jimmy.harvard.edu; li.david.wei@gmail.com

RESEARCH INTERESTS:

Develop computational technologies to understand the functions of coding and non-coding elements, especially in the context of human physiology and disease. I am interested in:

1. Method development for the design and analysis of functional screening (especially CRISPR/Cas9 knockout screening);
2. Transcriptome dynamics (alternative splicing and transcript structure changes) from high-throughput RNA sequencing data;
3. Functional analysis of coding and non-coding elements from integrating genomics big data.

EDUCATION:

- 2008.9-2012.6 Ph.D. in Computer Science and Engineering, University of California, Riverside, Riverside CA
- Dissertation: RNA-Seq Transcriptome Assembly: Sparsity, Bias Correction and Multiple Sample Comparison.
 - Committee: Tao Jiang (advisor), Stefano Lonardi, Marek Chrobak, Thomas Girke
- 2006.9-2008.6 Master in Computer Science, Tsinghua University, Beijing China
- Committee: Yannan Zhao (advisor), Zehong Yang
- 2002.9-2006.6 Bachelor in Computer Science, Tsinghua University, Beijing, China

CURRENT APPOINTMENT:

- 2012.9- Postdoctoral Research Fellow, Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health
- Advisor: Xiaole Shirley Liu

HONORS AND DISTINCTIONS:

- [1] APBC 2015 Best paper award, 2015
- [2] IEEE ICCABS conference Travel Award, 2012
- [3] Graduate Division Fellowship Award, University of California, Riverside, 2008-2009
- [4] Tsinghua University Student Art Group Fellowship, 2005
- [5] Tsinghua University Academic Career Fellowship, 2004
- [6] Director, Tsinghua University Military Band Freshman Division, 2006-2007
- [7] Deputy Director, Tsinghua University Military Band, 2004-2005
- [8] Freshman Excellence Scholarship, Tsinghua University, 2002
- [9] Highest ranking student in Guangdong Province College Entrance Exam (equivalent to SAT College Admission Exam; ranked 1st/~163,000)

PUBLICATIONS:

Submitted papers:

- [1] Tengfei Xiao*, **Wei Li***, Xiaoqing Wang, Han Xu, Qiu Wu, Peng Jiang, Jixin Yang, Teng Fei, Chongzhi Zang, Qi Liao, Jonathan Rennhack, Eran Andrechek, Rinath M. Jeselsohn, X. Shirley

Liu, Myles Brown. Drivers of Endocrine Resistance and Synthetic Lethal Vulnerabilities in Breast Cancer. Under revision in Nature (co-first author).

- [2] Teng Fei, Yiwen Chen, Tengfei Xiao, Laura Cato, Peng Zhang, **Wei Li**, Maura B. Cotter, Michaela Bowden, Lis T. Rosina, Massimo Loda, Shuang G. Zhao, Felix Y. Feng, Qiu Wu, Housheng Hansen He, X. Shirley Liu, Myles Brown. HNRNPL and its RNA Targets Regulate Prostate Cancer Growth. Submitted.

Journals:

- [3] Shiyong Zhu*, **Wei Li***, Jingze Liu, Chen-Hao Chen, Qi Liao, Han Xu, Tengfei Xiao, Zhongzheng Cao, Jingyu Peng, Pengfei Yuan, Myles Brown, Xiaole Shirley Liu & Wensheng Wei. CRISPR/Cas9-mediated genomic deletion screening for long non-coding RNAs using paired-gRNAs. In press, Nature Biotechnology. ***co-first author.**
- [4] Jian Ma, Johannes Köster, Qian Qin, Shengen Hu, **Wei Li**, Chenhao Chen, Qingyi Cao, Jinzeng Wang, Shenglin Mei, Qi Liu, Han Xu, and Xiaole Shirley Liu. CRISPR-DO for genome-wide CRISPR design and optimization. In press, Bioinformatics.
- [5] Shom Goel, Qi Wang, April C. Watt, Sara M. Tolaney, Deborah A. Dillon, **Wei Li**, Susanne Ramm, Adam C. Palmer, Haluk Yuzugullu, Kwok-Kin Wong, X. Shirley Liu, Piotr Sicinski, Eric P. Winer, Ian E. Krop, Jean J. Zhao. Overcoming Therapeutic Resistance in HER2-Positive Breast Cancers With CDK4/6 Inhibitors. Cancer Cell 2016, 29(3):255-269.
- [6] **Wei Li***, Johannes Köster*, Tengfei Xiao, Han Xu, Chen-Hao Chen, Jun S. Liu, Myles Brown, Xiaole S. Liu. Quality control, modeling and visualization of genome-wide CRISPR screens using MAGeCK-VISPR. Genome Biology 2015, 16:281.
- [7] Peng Jiang, Hongfang Wang, **Wei Li**, Chongzhi Zang, Bo Li, Yinling Wong, Cliff Meyer, Jun S. Liu, Jon Aster, X. Shirley Liu. Network analysis of gene essentiality in functional genomic screens. Genome Biology 2015, 16:239.
- [8] Masruba Tasnim, Shining Ma, Ei-Wen Yang, Tao Jiang* and **Wei Li***. Accurate Inference of Isoforms from Multiple Sample RNA-Seq Data. BMC Genomics 2015, 16 (S2):S15. **APBC 2015 Best Paper Award; *Corresponding author.**
- [9] Han Xu, Chen-Hao Chen, **Wei Li**, Cliff Meyer, Tengfei Xiao, Le Cong, Feng Zhang, Jun S. Liu, Myles Brown, X. Shirley Liu. Sequence determinants of improved CRISPR sgRNA design. Genome Research 2015, 25(8):1147-57.
- [10] Xuesong Zhao, Tatyana Ponomaryov, Kimberly J. Ornell, Pengcheng Zhou, Sukriti K. Dabral, Ekaterina Pak, **Wei Li**, Scott X. Atwood, Ramon J. Whitson, Anne Lynn S. Chang, Jiang Li, Anthony E. Oro, Jennifer A. Chan, Joseph F. Kelleher, and Rosalind A. Segal. RAS/MAPK activation drives resistance to Smo inhibition, metastasis and tumor evolution in Shh pathway-dependent tumors. Cancer Research 2015, 75(17):3623-35.
- [11] **Wei Li***, Han Xu*, Tengfei Xiao, Le Cong, Feng Zhang, Jun S. Liu, Myles Brown, X. Shirley Liu. MAGeCK enables robust identification of essential genes from genome-scale CRISPR-Cas9 knockout screens. Genome Biology 2014, 15:554.
- [12] Xiaoqi Zheng, Qian Zhao, Hua-Jun Wu, **Wei Li**, et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. Genome Biology 2014, 15:419.
- [13] **Wei Li** and Tao Jiang. Transcriptome Assembly and Isoform Expression Level Estimation from Biased RNA-Seq Reads. Bioinformatics (2012) 28(22):2914-2921.
- [14] Paul M. Ruegger, Elizabeth Bent, **Wei Li**, Daniel R. Jeske, Xinping Cui, Jonathan Braun, Tao Jiang, James Borneman. Improving Oligonucleotide Fingerprinting of rRNA Genes by Implementation of Polony Microarray Technology. Journal of Microbiological Methods (2012) 90(3):235-240.
- [15] **Wei Li**, Jianxing Feng, Tao Jiang. IsoLasso: A LASSO Regression Approach to RNA-Seq Based Transcriptome Assembly. Journal of Computational Biology, 18(11):1693-1707, Nov. 2011.
- [16] Jianxing Feng, **Wei Li**, Tao Jiang. Inference of isoforms from short sequence reads. Journal of Computational Biology, 18(3):305-321, Mar. 2011.

Conference proceedings:

- [17] Masruba Tasnim, Shining Ma, Ei-Wen Yang, Tao Jiang* and **Wei Li***. Accurate Inference of Isoforms from Multiple Sample RNA-Seq Data. 2015 Asian Pacific Bioinformatics Conference (APBC 2015). **APBC 2015 Best Paper Award**; *corresponding author; (the proceeding version of BMC Genomics 2015 paper).
- [18] **Wei Li**, *et al.* Workshop: Transcriptome assembly and isoform expression level estimation from biased RNA-Seq reads. IEEE 2nd Computational Advances in Bio and Medical Sciences (ICCABS 2012), pp. 1
- [19] **Wei Li**, *et al.* Workshop: Transcriptome assembly from RNA-Seq data: Objectives, algorithms and challenges. IEEE 1st Computational Advances in Bio and Medical Sciences (ICCABS 2011), pp. 271
- [20] **Wei Li**, Jianxing Feng, Tao Jiang. IsoLasso: A LASSO Regression Approach to RNA-Seq Based Transcriptome Assembly. Research in Computational Molecular Biology (RECOMB 2011). Lecture Notes in Computer Science, 6577:167-188; (the proceeding version of JCB 2011 paper).
- [21] **Wei Li**, Paul Ruegger, James Borneman, Tao Jiang. Polony Identification Using the EM Algorithm Based on a Gaussian Mixture Model. IEEE International Conference on Bioinformatics and BioEngineering (BIBE 2010), 220-225.
- [22] Jianxing Feng, **Wei Li**, Tao Jiang. Inference of isoforms from short sequence reads. Research in Computational Molecular Biology (RECOMB 2010). Lecture Notes on Computer Science, 6044:138-157. (the proceeding version of JCB 2011 paper).
- [23] **Wei Li**, *et al.* SVM feature selection and sample regression for Chinese medicine research. 2008 International Conference on Information and Antomation (ICIA 2008), pp. 1773.
- [24] **Wei Li**, *et al.* COX-2 activity prediction in Chinese medicine using neural network based ensemble learning methods. International Joint Conference on Neural Networks (IJCNN 2008), pp. 1853.

PRESENTATIONS:

- [1] *Computational Algorithms for CRISPR Screens and A Study of Breast Cancer Hormone Independence*. Workshop of CRISPR Congress 2016, Boston, MA
- [2] *Quality Control, Modeling and Visualization of Genome-wide CRISPR Screens Using MAGeCK-VISPR*. Frontiers in CRISPR Symposium: New Development and Applications of CRISPR-Cas9 for Genome-Editing. January 2016.
- [3] *Introduction To CRISPR Screens*. Seminar in Longwood Translational Medicine China Initiative. October 2015.
- [4] *Quality Control, Modeling and Visualization of Genome-wide CRISPR Screens Using MAGeCK-VISPR*. Cold Spring Harbor Laboratory symposium: "Genome Engineering: The CRISPR/Cas9 Revolution", September 2015 (oral presentation, acceptance rate 32%).
- [5] *Identification of cell-type specific essential genes using MAGeCK-VISPR*. Center for Functional Cancer Epigenetics Scientific Retreat, Dana-Farber Cancer Institute, August 2015.
- [6] *Quality Control, Modeling and Visualization of CRISPR Screens with MAGeCK-VISPR*. Genomics GetTogether Seminar, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, July 2015.
- [7] *CRISPR MAGeCK and Other Tricks*. Cell Circuits and Epigenomics Program Seminar. Broad Institute of Harvard and MIT. June 2015.
- [8] *Computational Analysis of Genome-Scale CRISPR-Cas9 Knockout Screening Experiments*. Center for Functional Cancer Epigenetics Seminar, Dana-Farber Cancer Institute, Mar. 2014.
- [9] *Transcriptome assembly from RNA-Seq data: Objectives, algorithms and challenges*. Computational Biology and Bioinformatics program, Yale University, Jun. 2012.
- [10] *Transcriptome assembly from RNA-Seq data: Objectives, algorithms and challenges*. Department of Stem Cell and Regenerative Biology, Harvard University, May 2012.

- [11] *Transcriptome Assembly and Isoform Expression Level Estimation from Biased RNA-Seq Reads*. IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS 2012). Las Vegas, NV. Mar. 2012.
- [12] *IsoLasso: A LASSO Regression Approach to RNA-Seq Based Transcriptome Assembly*. 15th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2011). Vancouver, Canada, Mar. 2012.
- [13] *Transcriptome assembly from RNA-Seq data: Objectives, algorithms and challenges*. IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS 2011). Orlando, FL. Feb. 2011.
- [14] *Polony Identification Using the EM Algorithm Based on a Gaussian Mixture Model*. 10th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2010). Philadelphia, PA, May 2010.
- [15] *COX-2 Activity Prediction in Chinese Medicine Using Neural Network Based Ensemble Learning Methods*. The 2008 International Joint Conference on Neural Networks (IJCNN 2008). Hong Kong, May 2008.

RESEARCH EXPERIENCE:

2012.9- Postdoctoral Research Fellow, Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health

Modeling the functions of coding and non-coding elements using CRISPR-Cas9 screens.

- Developed the MAGeCK (Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout) algorithm to process, analyze and visualize CRISPR/Cas9 knockout screening results; paper published in *Genome Biology* and received >10,000 visits; software had >7,000 downloads; see <http://mageck.sourceforge.net> for more details;
- Developed the MAGeCK-VISPR, a comprehensive quality control, modeling and visualization algorithm for CRISPR screens. The statistical model employs a negative-binomial framework that are optimized using the maximum likelihood approach. The MAGeCK-VISPR algorithm (<https://bitbucket.org/liulab/mageck-vispr>) is published in *Genome Biology*.
- Optimized gRNA pair design and analysis for long-range knockout of coding and non-coding elements, including long non-coding RNAs (lncRNAs) and enhancers in various systems (in collaboration with Myles Brown lab at DFCI and Wensheng Wei lab at Peking University). We designed a paired gRNA library targeting ~600 lncRNAs, performed screens on two cell lines. Almost all of the top lncRNA hits (found by MAGeCK) can be validated. The lncRNA screening paper was accepted by *Nature Biotechnology*;
- Developed algorithms to predict the efficiency of sgRNAs based on the sequence content (SSC & CRISPR-DO); manuscript published in *Genome Research and Bioinformatics*;
- Used genome-wide CRISPR screens to study the mechanisms and therapeutic solutions for breast cancer hormone independence (in collaboration with Myles Brown lab at DFCI). An integration of multiple CRISPR screening data (using MAGeCK and MAGeCK-VISPR) identified a critical tumor suppressor (CSK) whose loss directly leads to hormone independent cell growth. A secondary screening experiment identified synthetic lethal targets with CSK loss, and their small molecule inhibitors suppressed tumor growth in vitro and in vivo. Manuscript is under revision in *Nature*, and a related patent application is pending;

Building computational methodology for transcriptome dynamics and functional genomics from an integration of multiple genetic and epi-genetic sequencing data.

- Developed computational algorithms to identify somatic mutations and alternative splicing events from RNA-Seq normal-cancer pairs and in large cancer cohorts (rnaseqmut and ISP); ISP received **best paper award** at APBC 2015 conference;

- Studied somatic mutations and downstream gene expression changes from RNA-Seq data in PIK3CA driven breast tumor mouse models; identified key mutation events in recurrent breast tumors; top hits under validation;
- Studied the roles of PBRM1 loss-of-function mutations in renal cell carcinoma; Integrated gene expression, histone mark and PBRM1 ChIP-seq datasets to identify PBRM1 regulated gene networks, including target transcription factors and downstream genes.

2011.6-2011.9 Ph.D. research intern in Bioinformatics division, Novartis Institutes for Biomedical Research (NIBR), Emeryville, CA

- The 2nd Ph.D. intern since NIBR was established in Emeryville, CA.
- Helped NIBR set up RNA-Seq processing pipelines and performed large-scale RNA-Seq analysis over 4 TB data of >100 cancer patients.
- Performed computational analysis including detecting differentially expressed genes and transcripts, aberrant alternative splicing and fusion events and survival analysis.
- Identified several genes as potential drug targets in squamous cell carcinoma and breast cancer.

2008.9-2012.6 Ph.D. in Computer Science and Engineering, University of California, Riverside, Riverside CA

Developing assembly algorithms to reconstruct transcript structures from RNA-Seq data.

- Proposed IsoInfer to use a combinatorial search algorithm for de-novo transcript assembly from RNA-Seq read; developed IsoLasso to use graphical model with LASSO operator for transcript assembly and expression analysis; developed CEM to use quasi-multinomial to model different biases in RNA-Seq biases in transcript assembly.
- The first approach to use penalized least square approaches for the RNA-seq assembly problem; papers of IsoInfer, IsoLasso and CEM received **over 200** citations.

Building DNA identification algorithms to measure the interaction of probes and DNAs in a large-scale image.

- Proposed a mixture of 2-d Gaussian distributions to model the overlapping of probes, and using Expectation-Maximization (EM) algorithm to recursively identify the positions and intensities of overlapping probes.

2006.9-2008.6 Master in Computer Science, Tsinghua University, Beijing China

- Developed machine learning algorithms to predict the activity of several inhibitors of human inflammatory pathways found in a traditional Chinese medicine recipe ("Guizhitang", or 桂枝汤).
- Proposed two machine-learning models for the activity prediction, including a Support Vector Machine model and a neural-network based ensemble-learning model.
- Designed a MCMC self-localization algorithm for real-time localization in the soccer field on a mini LINUX-based soccer robot.

TEACHING EXPERIENCE:

2010 Teaching Assistant, University of California, Riverside

- The Theory of Automata and Formal Languages (CS150). Introduces fundamental concepts in automata theory and formal languages including grammar, finite automation, regular expression, formal language, pushdown automation and Turing machine. Lead the discussion seminar in the entire semester, once every week.

2011 Teaching Assistant, University of California, Riverside

- Machine Learning (CS229). Supervised machine learning method that emphasizes discriminative methods. Covers the areas of regression and classification. Topics include linear methods, instance-based learning, neural networks, kernel machines, and additive models.

2016 Guest lecturer, Harvard University

- Introduction to Computational Biology and Bioinformatics (STAT115). Introduction to next-generation sequencing, and basics for high-throughput RNA sequencing (RNA-seq)

GRANT APPLICATIONS:

Submitted applications:

- NIH Pathway to Independence award (K99/R00): Computational Characterization of Functional Elements using CRISPR Screens. Submitted Oct. 2015 and July 2016
- Department of Defense (DoD) prostate cancer research program (PCRP) postdoctoral training award: Submitted Nov. 2014
- Department of Defense (DoD) breast cancer research program (BCRP) postdoctoral training award: submitted Nov. 2013
- Other submitted applications: Burroughs Wellcome Fund (BWF) career awards at the scientific interface, Raymond Lavietes Foundation fellowship, Wong family award

Applications that I helped preparing:

- DoD PCRP PC140817 (Brown, Myles and Liu, X.S.): PCRP Synergistic Idea Development Award: Mechanism of Hormone Independence and Drug Resistance in Prostate Cancer (funded)
- NIH/NHGRI R01HG008728 (Brown, Myles and Liu, X.S.): Large-scale In Vivo Functional Characterization of the Human Cistrome (funded)
- NIH/NHGRI R01HG008927 (Liu, X.S.): Computational Methods for Genome-wide CRISPR/Cas9 Knockout Screens (funded)

MEMBERSHIPS AND SERVICE:

- Vice President, Harvard Medical School Chinese Scholars and Scientists Association (HMS-CSSA, <http://hms-cssa.org>), 2014-2015
 - Organized and hosted the Cancer Epigenetics Mini-symposium (2014), and organized the Single Cell Genomics Mini-symposium (2015);
 - Organized 2015 Harvard Chinese Life Science Annual Symposium, a whole-day seminar attracting 8 renown academic speakers and > 500 participants;
- Associate Director, Longwood Translational Medicine China Initiative (LTMCI, <http://www.lwtransmed.org/>)
 - Co-founder of the organization to facilitate the increasing collaboration among doctors and scientists from China and US. Organized over 40 seminars in 3 years
- Member of IEEE, ICSB.

PEER-REVIEWS:

Journals:

- Cell, Nature Biotechnology, PLoS Computational Biology, Nucleic Acid Research, Bioinformatics, Biostatistics, BMC Bioinformatics, BMC Genomics, BMC Cancer, Epigenetics and Chromatin, PLoS ONE, Journal of Computational Biology.

Conferences:

- ISMB/ECCB (2012/2013/2014), RECOMB (2013/2014), WABI (2011).

REFERENCE:

X. Shirley Liu

Professor of Biostatistics and Computational Biology, Harvard School of Public Health

Director, Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute

Associate member, Broad Institute

xslu@jimmy.harvard.edu

Myles Brown

Professor of Medicine, Harvard Medical School

Director, Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute

myles_brown@dfci.harvard.edu

Tao Jiang

Professor of Computer Science and Engineering

University of California, Riverside

jiang@cs.ucr.edu

Wensheng Wei

Professor, School of Life Sciences, Peking University, Beijing China

wsw@pku.edu.cn