

Cheng Li

January 2012

Objective: Research or faculty positions in computational biology and computational genomics

Contact

Associate Professor of Biostatistics

Department of Biostatistics, Harvard School of Public Health (HSPH)

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute (DFCI)

Center for Life Sciences, 3 Blackfan Circle, Boston, MA 02115

Homepage: <http://www.ChengLiLab.org/> or <http://www.biostat.harvard.edu/~cli/>

Phone: 617-632-3498

Email: cli@hsph.harvard.edu

Research Interests

Areas: computational biology, cancer genomics, bioinformatics, and applied statistical methods

How do genomic changes in the cell promote cancer initiation and progression, and can be used to predict patient response and prognosis? As biology and medicine increasingly become data-driven, we have unprecedented opportunity to integrate high-throughput microarray, sequencing, and genetic network data to study genomics, expression, and network changes in cancer cells towards novel drug targets and personalized medicine. To answer these questions, we develop novel statistical and computational methods as well as use existing ones such as hidden Markov models, Bayesian models and support vector machines. We are also dedicated to make user-friendly software that can be used by both data analysts and biologists, such as [dChIP](#) (cited 2100 times) and [ComBat](#) (cited 140 times).

Education

1991 – 1995 B.S. in Computer Science, Dept. of Mathematics, Beijing Normal University, China

1997 – 2001 Ph.D. in Statistics, Dept. of Statistics, University of California, Los Angeles

2001 – 2002 Research Fellow in Computational Biology, Dept. of Biostatistics, HSPH & DFCI

Positions and Services

1995 – 1997 Assistant Administrator, Computer Assisted Learning Lab, Beijing Normal University

2002 – 2008 Assistant Professor, Department of Biostatistics, HSPH & DFCI

2008 – Associate Professor, Department of Biostatistics, HSPH & DFCI

2009 – Guest Professor, Department of Bioinformatics, Tongji University, China

2006 – Associate Editor, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Journal Reviews

Annals of Applied Statistics, Biostatistics, PNAS, Nature Genetics, Bioinformatics, Nucleic Acids Research, Genome Biology, BMC Bioinformatics, BMC Genomics, BMC Genetics

Grant Reviews

- 2005.12 NIH RFA panel “Genomics of Transplantation Cooperative Research Program” (NIAID)
- 2007.04 NIH RFA panel “Research Centers of Excellence in Pediatric Nephrology” (NIDDK)
- 2007.04 NSF Electronic Proposal Review for “Arabidopsis 2010”
- 2009.06 NIH Challenge Grants RC1 Stage 1 reviewer
- 2009.10 NIH Biodata Management and Analysis (BDMA) Study Section
- 2010.06 NIH RFA panel “Metagenomic Evaluation of Oral Polymicrobial Disease” (NIDCR)

Research Support

- 2003 – 2004 PI, Friends of Dana-Farber Cancer Institute, "Developing analysis methods and software for SNP microarrays"
- 2004 – 2006 PI, Claudia Adams Barr Program in Cancer Research, "Developing analysis methods and software for oligonucleotide SNP microarrays"
- 2008 – 2011 PI, NIH R01GM077122 (NIGMS), “dChip: Analysis and Visualization of Oligonucleotide Expression and SNP Arrays”
- 2010 – 2011 PI, NIH R01GM077122-02S1 (NIGMS), “dChip: Analysis and Visualization of Oligonucleotide Expression and SNP Arrays”, Administrative supplement
- 2009 – 2011 Advisor, Claudia Adams Barr Program in Cancer Research, “Comparative Network Analysis of Human Cancers to Identify Conserved and Distinct Pathway Alterations” (PI: Parantu Shah of Cheng Li Lab)
- 2011 – 2016 Co-leader, Biostatistics and Bioinformatics Core of NIH P01CA155258 (NCI), “Integrative Oncogenomics of Multiple Myeloma” (PI: Nikhil Munshi)
- To resubmit PI, NIH R21CA152884 (NCI), “Analysis of transcription regulation altered by aneuploidy in cancer”, Review score: 13%

Teaching and Advising**Courses**

- Spring 2004 Co-instructor, BIO277 Computational Biology, HSPH
- Fall 2007 – 2009 Instructor, BIO237 Modern Statistical Computing Environments, HSPH
- Winter 2010 Co-instructor, BIO506 Introduction to Computational Biology, HSPH
- Fall 2010 – 2011 Instructor, BIO509 Introduction to Statistical Computing Environments, HSPH
- Fall 2010 – 2011 Instructor, BIO510 Programming I, HSPH
- Winter 2011 Instructor, BIO511 Programming II, HSPH

Ph.D. Students

- 2002 – 06 Yu Guo, “Statistical Issues in Microarray Data Analysis”
Current position: Manager of Biostatistics, BG Medicine
- 2004 – 07 Patrick Loerch, “Using Mixed Effects Models to Integrate High-Dimensional, Genomic Data and an Array-Based Analysis of the Evolution of Brain Aging”
Current position: Portfolio Strategy Team Lead, Merck
- 2010 – Norman Huang, “Integrative cancer genomics analysis”

Research Assistants (duration, current position)

- | | | |
|--------------|-----------|--|
| Sheng Zhong | 2002 – 03 | Assistant Professor, UIUC |
| Xuemin Fang | 2002 – 04 | McKinsey |
| Ming Lin | 2002 – 04 | Genentech |
| Ryung Kim | 2002 – 04 | Assistant Professor, Worcester Polytechnic Institute |
| Evan Johnson | 2004 – 05 | Assistant Professor, Boston University |
| Yunyu Zhang | 2004 – 05 | Novartis Institutes for BioMedical Research |
| Xi Deng | 2005 – 06 | |
| Jie Hu | 2007 – 08 | Ph.D student, University of Washington |
| Wai-Ki Yip | 2009 – 09 | Ph.D student, HSPH Biostatistics |
| Yulong Wang | 2010 – 11 | |
| Rui Zhao | 2010 – | |
| Mehmet Samur | 2010 – | |

Postdoctoral Fellows

- | | | |
|--------------|-----------|--|
| Igor Leykin | 2003 – 04 | |
| Yuhyun Park | 2004 – 05 | Boston Consulting |
| Samir Amin | 2008 – 11 | Ph.D student, Baylor College of Medicine |
| Aimin Yan | 2010 – 10 | AVEO Pharmaceuticals |
| Kai Wang | 2010 – 11 | Foundation Medicine |
| Zhenyu Yan | 2008 – | |
| Parantu Shah | 2009 – | |

Selected Publications (from 84)**Peer-reviewed Journals**

1. Yan Z, Shah PK, Amin SB, Samur MK, Huang N, Wang X, Misra V, Ji H, Gabuzda D, **Li C**. Transcription factor-miRNA feed-forward loops account for a large proportion of transcriptome changes and identify regulators in human cancers. *Nucleic Acids Res*. Revision submitted.
2. Huang N, Shah PK*, **Li C***. Lessons from a decade of integrating cancer copy number alterations with gene expression profiles. *Briefings in Bioinformatics*. 2012. Epub ahead of print.
3. Amin SB, Shah PK, Yan A, Adamia S, Minvielle S, Avet-Loiseau H, Munshi NC, **Li C**. The dChip survival analysis module for microarray data. *BMC Bioinformatics*. 2011, 12:72.
4. Cao Q, Zhou M, Wang X, Meyer CA, Zhang Y, Chen Z, **Li C***, Liu XS*. CaSNP: a database for interrogating copy number alterations of cancer genome from SNP array data. *Nucleic Acids Res*. 2011. 39(Database issue):D968-74.
5. Yan A*, Laird NM, **Li C***. Identifying rare variants using a Bayesian regression approach. *BMC Proceedings*. 2011, 5(Suppl 9):S99.
6. Chen M, Wang K, Zhang L, **Li C***, Yang Y*. The discovery of putative urine markers for the specific detection of prostate tumor by integrative mining of public genomic profiles. *PLoS One*. 2011;6(12):e28552.
7. Wang Y, Wu W, Negre NN, White KP, **Li C**, and Shah PK. Determinants of antigenicity and specificity in immune response for protein sequences. *BMC Bioinformatics*. 2011, 12:251
8. Ouillette P, Collins R, Shakhan S, Li J, Peres E, Kujawski L, Talpaz M, Kaminski M, **Li C**, Shedden K, and Malek SN. Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. *Blood*. 2011;118(11):3051-61.
9. Heinrichs S, **Li C**, Look AT. SNP array analysis in hematological malignancies: avoiding false discoveries. *Blood*. 2010; 115(21):4157-61.
10. Lipinski MM, Zheng B, Lu T, Yan Z, Py BF, Ng A, Xavier RJ, **Li C**, Yankner BA, Scherzer CR, Yuan J. Genome-wide analysis reveals mechanisms modulating autophagy in normal brain aging and in Alzheimer's disease. *Proc Natl Acad Sci*. 2010; 107(32):14164-9.
11. Parkin B, Erba H, Ouillette P, Roulston D, Purkayastha A, Karp J, Talpaz M, Kujawski L, Shakhan S, **Li C**, Shedden K, Malek SN. Acquired genomic copy number aberrations and survival in adult acute myelogenous leukemia. *Blood*. 2010; 116(23):4958-67.
12. **Li C**, Kim SW, Rai D, Bolla AR, Adhvaryu S, Kinney MC, Robetorye RS, Aguiar RC. Copy number abnormalities, MYC activity and the genetic fingerprint of normal B-cells mechanistically define the microRNA profile of DLBCL. *Blood*. 2009. 113(26):6681-90.
13. Avet-Loiseau H, **Li C**, Magrangeas, F, Gouraud, W, Charbonnel C, Harousseau JL, Attal M, Marit G, Mathiot C, Facon T, Moreau P, Anderson, KC, Campion L, Munshi NC, Minvielle S. Prediction of therapeutic efficacy based on gene expression profile in multiple myeloma In Press *J Clin Oncol*. 2009. 27(27):4585-90.
14. Shamma MA, Shmookler Reis RJ, Koley H, Batchu RB, **Li C**, Munshi NC. Dysfunctional homologous recombination mediates genomic instability and progression in myeloma. *Blood*. 2009; 113(10):2290-7.

15. Volchenbom SL, Li C, Li S, Attiyeh EF, Reynolds CP, Maris JM, Look AT, George RE. Comparison of Primary Neuroblastoma Tumors and Derivative Early-Passage Cell Lines Using Genome-Wide Single Nucleotide Polymorphism Array Analysis. *Cancer Res.* 2009; 69(10):4143-9.
16. Li J, Milbury CA, Li C, Makrigiorgos GM. Two-round coamplification at lower denaturation temperature-PCR (COLD-PCR)-based sanger sequencing identifies a novel spectrum of low-level mutations in lung adenocarcinoma. *Hum Mutat.* 2009. 30(11):1583-90.
17. Li C. Automating dChip: toward reproducible sharing of microarray data analysis. *BMC Bioinformatics.* 2008; 9:231.
18. Li C*, Beroukhim R, Weir BA, Winckler W, Garraway LA, Sellers WR, Meyerson M. Major copy proportion analysis of tumor samples using SNP arrays. *BMC Bioinformatics.* 2008; 9:204.
19. Loerch PM, Lu T, Dakin KA, Vann JM, Isaacs A, Geula C, Wang J, Pan Y, Gabuzda DH, Li C, Prolla TA, Yankner BA. Evolution of the aging brain transcriptome and synaptic regulation. *PLoS ONE.* 2008; 3(10):e3329.
20. Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, Hinkle G, Boehm JS, Beroukhim R, Weir BA, Mermel C, Barbie DA, Awad T, Zhou X, Nguyen T, Piquani B, Li C, Golub TR, Meyerson M, Hacohen N, Hahn WC, Lander ES, Sabatini DM, Root DE. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci.* 2008; 105:20380-5.
21. Johnson WE, Li C*, Rabinovic A. Adjusting batch effects in microarray expression data using Empirical Bayes methods. *Biostatistics.* 2007; 8(1): 118-127.
22. Park Y, Downing SR, Kim D, Hahn WC, Li C, Kantoff PW, Wei LJ. Simultaneous and exact interval estimates for the contrast of two groups based on an extremely high dimensional variable: Application to Mass Spec data. *Bioinformatics.* 2007; 23(12):1451-1458.
23. Vallat LD, Park Y, Li C, Gribben JG. Temporal genetic program following B-cell receptor cross-linking: altered balance between proliferation and death in healthy and malignant B cells. *Blood.* 2007; 109(9):3989-97.
24. Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway LA, Fox EA, Hochberg EP, Mellinghoff IK, Hofer MD, Descazeaud A, Rubin MA, Meyerson M, Wong WH, Sellers WR*, Li C*. Inferring Loss-of-Heterozygosity From Tumor-only Samples Using High-Density Oligonucleotide SNP Arrays. *PLoS Computational Biology.* 2006; 2(5): e41. 1-10.
25. Walker BA, Leone PE, Jenner MW, Li C, Gonzalez D, Johnson DC, Ross FM, Davies FE, Morgan GJ. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. *Blood.* 2006;108: 1733-1743.
26. Leykin I, Hao K, Cheng J, Meyer N, Pollak MR, Smith RJ, Wong WH, Rosenow C*, Li C*. Comparative linkage analysis and visualization of high-density oligonucleotide SNP array data. *BMC Genetics.* 2005; 6:7. 1-16.
27. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML, Fisher DE, Sellers WR. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature.* 2005; 436(7047):117-22.
28. LaFramboise T, Weir BA, Zhao X, Beroukhim R, Li C, Harrington D, Sellers WR, Meyerson M. Allele-Specific Amplification in Cancer Revealed by SNP Array Analysis. *PLoS Comput Biol.* 2005; 1(6): e65.

29. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH*, Li C*. dChipSNP: Significance Curve and Clustering of SNP-Array-Based Loss-of-Heterozygosity Data. *Bioinformatics*. 2004; 20: 1233-1240.
30. Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, Yankner BA. Gene regulation and DNA damage in the ageing human brain. *Nature*. 2004; 429: 883–891.
31. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, Girard L, Minna J, Christiani D, Leo C, Gray JW, Sellers WR, Meyerson M. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Research*. 2004; 64:3060-3071.
32. Janne PA, Li C, Zhao X, Girard L, Chen TH, Minna J, Christiani DC, Johnson BE, Meyerson M. High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene*. 2004; 23(15):2716-26.
33. Zhong S, Li C*, Wong WH*. (2003) ChipInfo: Software for Extracting Gene Annotation and Gene Ontology Information for Microarray Analysis. *Nucleic Acids Research*. 2003; 31:3483-3486.
34. Lieberfarb ME, Lin M, Lechpammer M, Li C, Tanenbaum DM, Febbo PG, Wright RL, Shim J, Kantoff PW, Loda M, Meyerson M, Sellers WR. Genome-wide Loss of Heterozygosity Analysis from Laser Capture Microdissected Prostate Cancer Using Single Nucleotide Polymorphic Allele (SNP) Arrays and a Novel Bioinformatics Platform dChipSNP. *Cancer Research*. 2003; 63(16): 4781-4785.
35. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl. Acad. Sci.* 2001; 98: 31-36.
36. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biology*. 2001; 2(8): research0032.1-0032.11
37. Schadt EE, Li C, Ellis B and Wong WH. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *Journal of Cellular Biochemistry*. 2001; Supplement 37, 120-125.
38. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, Haroutunian V and Fienberg AA. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc. Natl. Acad. Sci.* 2001; Vol. 98, 4746-4751.
39. Irizarry K, Kustanovich V, Li C, Brown N, Nelson S, Wong W and Lee CJ. Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences, *Nature Genetics*. 2000; 26:233-236.

(* Corresponding authors)

Book Chapters

40. Yip WK, Amin SB, and Li C. A Survey of Classification Techniques for Microarray Data Analysis. In *Handbook of Computational Statistics: Statistical Bioinformatics*. Edited by Henry Horng-Shing Lu, Bernhard Schölkopf, and Hongyu Zhao. Springer. 2011. 193-224.
41. Li C, Amin S. Analysis of cancer genome alterations using single nucleotide polymorphism (SNP) microarrays. In *New developments in biostatistics and bioinformatics*. Edited by Jianqing Fan, Xihong Lin and Jun Liu. World Scientific Publishing Co. 2009; 209-237.

42. Johnson WE and **Li C**. Adjusting batch effects in microarray experiments with small sample size using Empirical Bayes methods. In *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*. Edited by Andreas Scherer. Wiley Series in Probability and Statistics. 2009; 113-129.
43. Fang X, **Li C**, Wong WH. Low-level analysis of oligonucleotide expression arrays. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. John Wiley & Sons. 2005.
44. **Li C**, Wong WH. DNA-Chip Analyzer (dChip). In *The analysis of gene expression data: methods and software*. Edited by Parmigiani G, Garrett ES, Irizarry R and Zeger SL. Springer, New York. 2003; 120-141.

Seminar and Conference Talks

- 2006.04 St. Jude Hospital, Hartwell Center for Bioinformatics and Biotechnology, Memphis
- 2006.05 University of California, Riverside, Department of Statistics
- 2006.06 International Bioinformatics Workshop (IBW). Central South University, China
- 2006.09 Affymetrix 500K SNP Array Data Analysis Workshop, Boston
- 2006.10 University of Michigan School of Public Health, Department of Biostatistics
- 2006.11 Univ. of Rochester Medical Center, Dept. of Biostatistics and Computational Biology
- 2007.03 Brown University, Center for Statistical Sciences
- 2007.06 International Conference on Bioinformatics, Zhejiang University, China
- 2007.09 MD Anderson Cancer Center, Department of Bioinformatics
- 2008.02 Biogen Idec, Cambridge, MA
- 2008.03 University of Texas Health Science Center, San Antonio Cancer Institute
- 2008.09 Michigan State University, Science at the Edge seminar series
- 2008.12 Tongji University, School of Life Sciences, China
- 2009.06 International Bioinformatics Workshop (IBW), Suzhou University, China
- 2009.10 Yale School of Public Health, Division of Biostatistics
- 2009.12 Emory University, Department of Biostatistics and Bioinformatics
- 2010.05 Harvard Chinese Life Science Annual Research Symposium
- 2010.12 Conference on Bio-Inspired Models (Bionetics 2010), Keynote speaker.
- 2011.02 Yale University, Department of Statistics
- 2011.04 Tongji University Genomics Symposium, Shanghai, China