

Zhaohui (Steve) Qin

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University
Atlanta, GA 30322
Tel: (404) 712-9576
Fax: (404) 727-1370
Email: zhaohui.qin@emory.edu
URL: <http://userwww.service.emory.edu/~zqin4/>

Current position

Associate Professor	Aug 2010 – Present
Department of Biostatistics and Bioinformatics	
Rollins School of Public Health	
Emory University	Atlanta, GA

Additional affiliation

Center for Comprehensive Informative, Emory University
Graduate Program in Population Biology, Ecology and Evolution, Emory University (pending)
Michigan Center for Translational Pathology, University of Michigan

Education

B.S. in Probability and Statistics	Sep 1990 – Jun 1994
Peking University	Beijing, China
M.S. in Statistics	Sep 1995 – Jun 1997
University of Michigan	Ann Arbor, MI
Ph.D. in Statistics	Sep 1995 – Jun 2000
University of Michigan	Ann Arbor, MI

Professional experience

Graduate Student Instructor	Sep 1995 – May 1999
Department of Statistics	
University of Michigan	Ann Arbor, MI

Graduate Student Research Assistant Department of Statistics and Management Science University of Michigan Business School	Sep 1998 – May 1999 Ann Arbor, MI
Visiting Research Assistant Department of Statistics Stanford University	Aug 1999 – Sep 1999 Stanford, CA
Graduate Student Research Assistant Center for Statistics Consulting and Research University of Michigan	Sep 1999 – Aug 2000 Ann Arbor, MI
Postdoctoral fellow Department of Statistics, Advisor: Jun S. Liu Harvard University	Sep 2000 – Aug 2003 Cambridge, MA
Assistant Professor Center for Statistical Genetics, Department of Biostatistics University of Michigan	Sep 2003 – Jul 2010 Ann Arbor, MI
Core member of long program Institute of Pure and Applied Mathematics Workshop on Mathematical and Computational Approaches on High Throughput Genomics University of California at Los Angeles,	Sep 2011 – Dec 2011 Los Angeles, CA

Research Funding

Active PI grant:

R01HG005119-01 “*Model-based Methods for Analyzing ChIP Sequencing Data*”
PI: Qin
07/22/09 – 06/30/13

R21HG004751-01A2 “*Statistical Methods for Analyzing Resequencing Data*”
PI: Qin
09/21/10 – 07/31/12

DMS1000617 “*Spatial Model-Based Methods for RNA-seq Data Analysis*”
Multiple PIs: Zhu (Purdue University), Qin
10/01/10 – 09/30/13

Completed PI or co-PI grants:

University of Michigan Rackham Graduate School Interdisciplinary workshop. “*New Statistical Methods in Molecular Biology*”. (PD: Qin) 2004-2008.

2006 University of Michigan Rackham Graduate School Faculty Grants “*Using Model Based Clustering Method to Infer Expression Modules in Cancer*”. (PI: Qin).

University of Michigan Center for Computational Medicine and Biology 2006 pilot grant “*Joint Modeling of OMICS Data to Prioritize Drug Target Genes in B. anthracis*”. (Multiple PIs: Qin, Bergman).

2007 University of Michigan Rackham Graduate School Spring/Summer Research Grants. “*Developing Efficient Query Algorithm to Detect Functionally Related Genes from Large Set of Microarray Gene Expression Data*”. (PI: Qin).

2007 University of Michigan Office of Vice President for Research Faculty Grant “*Model-based Clustering of Mammalian Conserved DNA Regulatory Elements*”. (PI: Qin).

The Genetics and Genomics 2008 Pilot Feasibility Grant. University of Michigan Center for Genomics Health and Medicine “*Targeted Resequencing Study of Seven New Psoriasis Candidate Genes*”. (PI: Qin).

Career Development Research Project. The University of Michigan Comprehensive Cancer Center Prostate SPORE “*Model-based Methods to Analyze Massive Parallel Sequencing Data with Application to Prostate Cancer Progression*”. (PI: Qin) 11/01/08 – 10/31/10

The Genetics and Genomics 2008 Pilot Feasibility Grant. University of Michigan Center for Genomics Health and Medicine “*Use of ChIP-Seq to identify target genes in a novel, regulated transcriptional repressor-activator relationship*”. (PI: Schwartz, Qin).

The Genetics and Genomics 2010 Pilot Feasibility Grant. University of Michigan Center for Genomics Health and Medicine “*ChIP-Seq Analysis of Transcriptional Coactivator PGC-1 α in the Control of Circadian Energy Metabolism in the Liver*”. (PI: Lin, Qin).

Completed co-I grants (selected)

Department of Defense W81XWH-08-1-0110 “*A Search for Gene Fusions/ Translocations in Breast Cancer*” (PI: Chinnaiyan).

Role: Co-investigator.

American Diabetes Association 7-09-BS-168 “*Role of transcriptional repressor Bcl-6 in GH-induced insulin resistance*”. (PI: Schwartz).

Role: Co-investigator.

Peer reviewed journal and conference publications

* trainee.

joint first author.

Chen L, **Qin ZS**, Liu JS. (2001) Exploring Hybrid Monte Carlo in Bayesian Computation. *Bayesian Methods with Application to Science, Policy and Official Statistics*. 71-80.

Qin ZS, Liu JS. (2001) Multi-Point Metropolis Method with Application to Hybrid Monte Carlo. *J. Comp. Phys.* **172** 827-840.

#Niu T, #**Qin ZS**, Xu X, Liu JS. (2002) Bayesian Haplotype Inference for Multiple Linked Single Nucleotide Polymorphisms. *Am. J. Hum. Genet.* **70** 157-169.

#**Qin ZS**, #Niu T, Liu JS. (2002) Partition-Ligation EM Algorithm for Haplotype Inference with Single Nucleotide Polymorphisms. *Am. J. Hum. Genet.* **71** 1242-1247.

Qin ZS, McCue LA, Thompson W, Mayerhoffer L, Lawrence CE, Liu JS. (2003) Identification of Co-regulated Genes Through Bayesian Clustering of Predicted Regulatory Binding Sites. *Nat Biotechnol.* **21** 435-439.

Lu X, Zhang W, **Qin ZS**, Kwast KE, Liu JS. (2004) Statistical Resynchronization and Bayesian Detection of Periodically Expressed Genes. *Nucleic Acids Res.* **32** 447-455.

Kang H, **Qin ZS**, Niu T, Liu JS. (2004) Incorporating Genotyping Uncertainty in Haplotype Inference for Single-Nucleotide Polymorphisms. *Am. J. Hum. Genet.* **74** 495-510.

Zhang K, **Qin ZS**, Liu JS, Chen T, Waterman MS, Sun F. (2004) Haplotype Block partitioning and Tag SNP Selection Using Genotype Data and Their Applications to Association Studies. *Genome Res.* **14** 908-916.

Zhang K, **Qin ZS**, Chen T, Liu JS, Waterman MS, Sun F. (2004) HapBlock: Haplotype Block Partitioning and Tag SNP Selection Software Using a Set of Dynamic Programming Algorithms. *Bioinformatics.* **21** 131-134.

Niu T, Lu X, Kang H, **Qin ZS**, Liu JS. (2004) Haplotype Inference and Its Application in Linkage Disequilibrium Mapping. Computational Methods in SNPs and Haplotype Inference. DIMACS/RECOMB Satellite Workshop, Piscataway, NJ, USA, November 21-22, 2002, Editors: Sorin Istrail, Michael Waterman, Andrew Clark. Springer-Verlag. 48-61.

Zhang K, Chen T, Waterman MS, **Qin ZS**, Liu JS, Sun F. (2004) Dynamic Programming Algorithms for Haplotype Block Partitioning and Tag SNP Selection Using Haplotype Data or Genotype Data DIMACS/RECOMB Satellite Workshop, Piscataway, NJ, USA, November 21-22, 2002. Editors: Sorin Istrail, Michael Waterman, Andrew Clark. Springer-Verlag. 96-112.

*Zhu D, **Qin ZS** (2005) Structure Comparison of Metabolic Networks in Selected Single Cell Organisms. *BMC Bioinformatics.* **6** 8.

Cluster 17 Collaboration (2005) Fine Mapping of the Psoriasis Susceptibility Gene PSORS1: A Reassessment of Risk Associated With a Putative Risk Haplotype Lacking HLA-Cw6. *J Invest Dermatol.* **124** 921-930.

*Zhu D, Hero AO, **Qin ZS**, Swaroop A. (2005) High Throughput Screening of Co-expressed Gene Pairs with Controlled False Discovery Rate (FDR) and Minimum Acceptable Strength (MAS). *J. Comput. Biol.* **12** 1029-1045. (Publication resulted from summer rotation projects).

Nistor I, Nair RP, Stuart P, Abecasis GR, Hiremagalore R, Thompson RA, Jenisch S, Weichenthal M, Abecasis GR, **Qin ZS**, Christophers E, Lim HW, Voorhees JJ, Elder JT. (2005) Protein Tyrosine Phosphatase Gene PTPN22 Polymorphism in Psoriasis: Lack of Evidence for Association. *J Invest Dermatol.* **125** 395-396.

Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P; International HapMap Consortium. A Haplotype Map of the Human Genome. *Nature.* **437** 1299-1320.

Stuart P, Nair RP, Abecasis GR, Nistor I, Hiremagalore R, Chia NV, **Qin ZS**, Thompson RA, Jenisch S, Weichenthal M, Janiga J, Lim HW, Christophers E, Voorhees JJ, Elder JT. (2006) Analysis of RUNX1 Binding Site and RAPTOR Polymorphisms in Psoriasis: No Evidence for Association Despite Adequate Power and Evidence for Linkage. *J Med Genet.* **43** 12-17.

Qin ZS, *Gopalakrishnan S, Abecasis GR. (2006) An Efficient Comprehensive Search Algorithm for TagSNP Selection Using Linkage Disequilibrium Criteria. *Bioinformatics.* **22** 220-225.

*Ulitz PJ, Zhu J, **Qin ZS**, Andrews PC. (2006) Improved Classification of Mass Spectrometry Database Search Results Using Newer Machine Learning Approaches. *Mol Cell Proteomics.* **5** 497-509.

Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, Halperin E, Lin S, **Qin ZS**, Munro HM, Abecasis GR, Donnelly P; International HapMap Consortium. (2006) A Comparison of Phasing Algorithms for Trios and Unrelated Individuals. *Am J Hum Genet.* **78** 437-450.

*Gopalakrishnan S, **Qin ZS**. (2006) TagSNP Selection Based on Pairwise LD Criteria and Power Analysis in Association Studies. *Pac Symp Biocomput.* **11** 511-522.

Qin ZS. (2006) Clustering Microarray Gene Expression Data Using Weighted Chinese Restaurant Process. *Bioinformatics.* **22** 1988-1997.

Bergman NH, Passalacqua KD, Hanna PC, **Qin ZS**. (2007) Operon Prediction in Sequenced Bacterial Genomes without Experimental Information. *Appl Environ Microbiol.* **73** 846-854.

Xiang Z, **Qin ZS**, He Y. (2007) CRCView: A Web Server for Analyzing and Visualizing Microarray Gene Expression Data Using Model-based Clustering. *Bioinformatics.* **23** 1843-1845.

*Ma J, **Qin ZS**. (2007) Different Normalization Strategies for Microarray Gene Expression Traits Affect the Heritability Estimation. *BMC Proc.* **1 (Suppl 1)** S154.

Bommer GT, Gerin I, Feng Y, Kaczorowski AJ, Kuick R, Love RE, Zhai Y, Giordano TJ, **Qin ZS**, Moore BB, MacDougald OA, Cho KR, Fearon ER. (2007) p53-mediated Activation of miRNA34 Candidate Tumor-suppressor Genes. *Curr Biol.* **17** 1298-1307.

The International HapMap Consortium. (2007) A Second Generation Human Haplotype Map of Over 3.1 Million SNPs. *Nature.* **449** 851-861.

Sabeti PC and the International HapMap Consortium. (2007) Genome-wide Detection and Characterization of Positive Selection in Human Populations. *Nature.* **449** 913-918.

Qin ZS, Zhu J, Ulintz PJ, Andrews PC. (2008) Machine Learning Approaches for Peptide Identification based on Mass Spectrometry Database Search Results. *Introduction to Machine Learning and Bioinformatics*. Editors: Sushmita Mitra, Sujay Datta, Theodore Perkins, George Michailidis. CRC Press.

Zandi PP, Zöllner S, Avramopoulos D, Willour VL, Chen Y, **Qin ZS**, Burmeister M, Miao K, *Gopalakrishnan S, McEachin R, Potash JB, DePaulo JR, McInnis MG. (2008) Family-based SNP Association Study on 8q24 in Bipolar Disorder. *Am J Med Genet B: Neuropsychiatr Genet.* **147B** 612-618.

Sandstedt SA, Zhang L, Patel M, McCrea KM, **Qin ZS**, Marrs CF, Gilsdorf JR. (2008) Comparison of Laboratory-based and Phylogenetic Methods to Distinguish Between *Haemophilus influenzae* and *H. haemolyticus*. *J Microbiol Methods.* **75** 369-371.

*Zhou J, **Qin ZS**, Kim S, Wang Z, Yu M, Chien JY, Lucksiri A, Li L. (2009) A New Probabilistic Rule for Drug-Drug Interaction Prediction. *J Pharmacokinet and Pharmacodyn.* **36** 1-18.

*Hu M, **Qin ZS**. (2009) Apply Bayesian Model Selection to Better Query Large Scale Microarray Gene Expression Datasets. *PLoS ONE.* **4** e4495.

Moroi SE, Raouf DA, Reed DM, Zöllner S, **Qin ZS**, Richards JE. (2009) Progress toward Personalized Medicine for Glaucoma. *Expert Review of Ophthalmology.* **4** 145-161.

*Zhou J, **Qin ZS**, Quinney KS, Kim S Wang Z, Hall S, Li L. (2009) Drug-Drug Interactions Prediction Assessment. *J Biopharm Stat.* **19** 641-657.

*Chen Y, Lin G, Huo JS, Barney D, Wang ZN, Livshiz T, States DJ, **Qin ZS**, Schwartz J. (2009) Computational and Functional Analysis of Growth Hormone-regulated Genes Identifies the Transcriptional Repressor Bcl6 as a Participant in GH-regulated Transcription. *Endocrinology.* **150** 1645-1654.

Gudjonsson JE, Ding J, Li X, Nair R, Tejasvi T, **Qin ZS**, Ghosh D, Aphale A, Gumucio DL, Voorhees JJ, Abecasis G, Elder JT. (2009) Global Gene Expression Analysis Reveals Evidence for Decreased Lipid Biosynthesis and Increased Innate Immunity in Uninvolved Psoriatic Skin. *J Invest Dermatol.* **129** 2795-2804.

*Choi H, Nesvizhskii A, Ghosh D, **Qin ZS**. (2009) Hierarchical Hidden Markov Model with Application to Joint Analysis of ChIP-chip and ChIP-seq Data. *Bioinformatics* **25** 1715-1721.

*Choi H, **Qin ZS**, Ghosh D. (2010) Combined Genomic Profiles of DNA Copy Number and mRNA Expression Data and Mixture Modeling Framework. *J. Comput. Biol.* **17** 121-137.

*Hu M, Yu J, Taylor, JM, Chinnaiyan AM, **Qin ZS**. (2010) On the Detection and Refinement of Transcription Factor Binding Sites Using ChIP-Seq Data. *Nucleic Acids Res.* **38** 2154-2167.

Yu J, Yu J, Mani R, Cao X, Cao Q, Brenner CJ, Cao X, Wang GX, Wu L, Li J, *Hu M, Gong Y, Cheng H, Laxman B, Vellaichamy A, Shankar S, Li Y, Dhanasekaran SM, Morey R, Barrette T, Lonigro RJ, Tomlins SA, Varambally S, **Qin ZS**, Chinnaiyan, AM. (2010) An Integrated Network of Androgen Receptor, Polycomb and TMPRSS2-ERG Gene Fusion in Prostate Cancer Progression. *Cancer Cell* **17** 443-454. (From the cover with preview article in the same issue: Chen and Sawyers 415-416)

Breitkreutz A, Choi H, Sharom J, Boucher L, Neduva V, Larsen B, Lin Z, Breitkreutz B, Stark C, Liu G, Ahn J, Dewar-Darch D, **Qin ZS**, Pawson T, Gingras A, Nesvizhskii AI, Tyers M. (2010) A global protein kinase and phosphatase interaction network in yeast. *Science* **328** 1043-1046 (preview article in the same issue Levy et al. 983-984).

Qin ZS, Yu J, *Shen J, Maher CA, *Hu M, Kalyana-Sundaram S, Yu J, Chinnaiyan AM. (2010) HPeak: An HMM-based Algorithm for Defining Read-enriched Regions in ChIP-Seq Data. *BMC Bioinformatics.* **11** 369.

Ghosh D, **Qin ZS**. (2010) Statistical Issues in the Analysis of ChIP-Seq and RNA-Seq Data. *Genes* **1** 317-334.

Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, **Qin ZS**, Tyers M, Gingras AC, Nesvizhskii AI. (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods.* **8** 70-73.

Cui TX, Lin G, LaPensee CR, Calinescu A, Rathore M, Streeter C, Pwien-Pilipuk G, Lanning N, Jin H, Carter-Su C, **Qin ZS**, Schwartz J. (2011) C/EBP β mediates Growth Hormone-regulated expression of multiple target genes. *Molecular Endocrinology* **25** 681-693.

Kim JH, Dhanasekaran SM, Prensner JR, Cao X, Robinson D, Kalyana-Sundaram S, Shankar S, Jing X, Iyer M, *Hu M, Sam L, Grasso C, Maher CA, Palanisamy N, Mehra R, Huang C, Siddiqui J, Yu Y, **Qin ZS**, Chinnaiyan AM. (2011) Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res.* **21** 1028-1041.

Cao Q, Mani R, Ateeq B, Dhanasekaran SM, Asangani IA, Prensner JR, Kim JH, Brenner JC, Jing X, Cao X, Wang R, Li Y, Dahiya A, Wang L, Pandhi M, Lonigro RJ, Wu Y, Tomlins SA, Palanisamy N, **Qin ZS**, Yu J, Maher CA, Varambally S, Chinnaiyan AM. (2011) Coordinated

Regulation of Polycomb Group Complexes through microRNAs in Cancer. *Cancer Cell*. **20** 187-199.

Submitted

Hu M, Zhu Y, Taylor JMG, Liu JS, **Qin ZS** (2011). Using Poisson mixed-effects model to quantify exon-level gene expression in RNA-seq.

Other publications

Qin ZS, Damien P, Walker S. (2003) Scale Mixture Models with Applications to Bayesian Inference. *The Monte Carlo Methods in the Physical Sciences: Celebrating the 50th Anniversary of the Metropolis Algorithm*, Los Alamos, New Mexico 2003. Gubernatis JE (Ed) American Institute of Physics. 394-395.

Qin ZS. (2008) Book Review: *Data Analysis and Graphics Using R: An Example-Based Approach*, Second Edition. John Maindonald and John Braun *Brief Bioinform.* **9** 258-259.

Oral presentations (within 6 years)

Invited/selected conference presentations

An Efficient Comprehensive TagSNP Selection Algorithm. *Genomics Study and the HapMap*. March 15-18, 2005. Oxford, UK.

An Efficient Comprehensive Search Algorithm for TagSNP Selection Using Linkage Disequilibrium Criteria. *American Society of Human Genetics Annual Meeting*. October 25-29, 2005. Salt Lake City, Utah.

Processing, Normalization of Gene Expression Traits and Their Effect on Analysis. *GAW 15*. November 11-15, 2006. *Genetics Analysis Workshop 15*. Tampa, Florida (co-presenter of group work).

Grouping Genome-wide Sets of Mammalian Conserved Elements with Scalable Bayesian Clustering. *Indy Regional Bioinformatics Conference 2007*. Indianapolis, IN, May 31 – June 2, 2007.

A Set of Model-based Knowledge Discovery Tools for Analyzing Large Microarray Gene Expression Datasets. *First Annual Midwest Symposium on Computational Biology and Bioinformatics*. Evanston, IL, October 5, 2007.

Bayesian Methods for Haplotype Inference from Unrelated Individuals. *Haplotype Analysis of Population and Pedigree Data in Association Studies*. Birmingham, AL May 7-8, 2008.

Query Large Scale Microarray Gene Expression Datasets Using Bayesian Model-based Method with Variable Selection. *Statistical Genetics Workshop at Michigan State University*. Lansing, MI. May 14, 2008.

Bayesian Model-based Methods for Analyzing ChIP Sequencing Data. *ENAR 2010 Spring meeting*. New Orleans, LA. March 20-24, 2010.

Bayesian Model-based Methods for Analyzing RNA-Seq Data. *ENAR 2011* Spring meeting. Miami, FL. March 20-23, 2011.

Do chromosomes have “consensus” 3D structures? -- Statistical clues from Hi-C data. *Cold Spring Harbor Asia Conferences series 2011: High Throughput Biology*. Suzhou, China, April 19-23, 2011.

Bayesian Model-based Methods for Analyzing RNA-seq Data. *Sixth International Conference on Dynamic Systems and Applications*. Atlanta, GA. May, 25-28, 2011

Bayesian Model-based Methods for Analyzing RNA-seq Data. 2011 International Research Symposium on Frontiers of Statistics. Hefei, China. July 13-18, 2011.

Invited departmental seminars

Operon Prediction in Newly Sequenced Genomes Using HMM. ASA Central Indiana Chapter. April 19, 2005. Indianapolis, IN.

Model-based Methods for Mining Large Scale Genomic Data, Department of Statistics, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, November 8, 2007.

HPeak: A Model-Based Algorithm for Defining Read-Enriched regions. Department of Human Genetics, University of Michigan. Next-Generation Sequencing Seminar Series. Ann Arbor, MI. July 14, 2008.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology. University of Pennsylvania. Philadelphia, PA. March 2, 2010.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. Department of Computer Science, University of New Orleans. New Orleans, LA. March 22, 2010.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. Institute of Computational Biology, Weill Cornell Medical College. New York, NY. March 25, 2010.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. Statistical Bioinformatics Center, Department of Statistics, Purdue University. West Lafayette, IN. March 30, 2010.

Bayesian model-based methods for analyzing ChIP sequencing data. Computational Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA. October 14, 2010.

Statistical Modeling of RNA-seq Data. Division of Biostatistics, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL. October 25, 2010.

Bayesian model-based methods for analyzing ChIP sequencing data. Center for Bioinformatics and Computational Genomics, School of Biology, Georgia Institute of Technology, Atlanta, GA. November 23, 2010.

Bayesian model-based methods for analyzing RNA-Seq data. Department of Mathematics and Statistics, Georgia State University, Atlanta, GA. March 18, 2011.

Using model-based methods to quantify exon-level gene expression from RNA-seq data. Department of Biomathematics, UCLA, Los Angeles, CA. September 29, 2011.

Model-based methods for analyzing NGS data. Department of Statistics, Yale University, New Haven, CT. October 17, 2011.

Model-based methods for analyzing NGS data. Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA. November 3, 2011.

Poster presentations

MotifOrganizer: A Scalable Multi-stage Model-based Clustering Approach for Grouping Conserved Non-coding Elements in Mammalian Genomes. Presented at the Conference on Emerging Design and Analysis Issues in genomics Studies in Population Sciences. Boston, MA. October 11-12, 2007.

Molecular Cross-talk of Androgen Receptor and TMPRSS2:ERG Gene Fusion Product in Prostate Cancer, co-presented with Dr. Jindan Yu. 2008 Cold Spring Harbor Meeting on System Biology: Global Regulation of Gene Expression. Cold Spring Harbor, NY. March 27-30, 2008.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. Pacific Symposium on Biocomputing (PSB) 2010. Hawaii, HI. January 4-8, 2010.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. 2010 Cold Spring Harbor Meeting on System Biology: Global Regulation of Gene Expression. Cold Spring Harbor, NY. March 23-27, 2010.

Detection and Refinement of Transcription Factor Binding Sites Using Hybrid Monte Carlo Method. Frontiers in Mathematical Biology: NSF-NIH PIs Meeting 2010. College Park, MD. April 26-27, 2010

Software developed

- **EM-DeCODER** (haplotype phasing using EM algorithm).
- **HAPLOTYPYER** (haplotype phasing using Partition-Ligation and Gibbs Sampler).
- **PL-EM** (haplotype phasing using Partition-Ligation and EM algorithm).
- **BMC** (putative binding motif clustering using Gibbs Sampler).
- **FESTA** (tagSNP selection based on pairwise LD) With Shyam Gopalakrishnan.
- **tripleM** (PL-EM for phasing trios).
- **CRC** (model-based clustering algorithm for gene expression microarray data).
- **OperonHMM** (a hidden Markov model for predicting operon structure using only sequence data). Co-developed with Dr. Nick Bergman.
- **CRCView** (point-and-click web server for clustering, visualizing and interpreting microarray gene expression data). Co-developed with Zuoshuang Xiang and Yongqun He.
- **BEST** (query large scale microarray compendium datasets using model-based Bayesian approach with variable selection). Developed by Ming Hu.
- **HPeak** (A hidden Markov model-based algorithm for calling ChIP-enriched peaks using ChIP-seq data).
- **ChIP-meta** (hierarchical hidden Markov model-based algorithm for inferring ChIP-enriched regions using both ChIP-seq and ChIP-chip data). Developed by Hyung Won Choi.
- **HMS** (a hybrid Monte Carlo-based *de novo* motif discovery tool designed specifically for analyzing ChIP-Seq data). Developed by Ming Hu.

Professional services

Editorial board

Plant Cell 2011-present

Review editor board

Frontier in Statistical Genetics and Methodology (a specialty of Frontiers in Genetics) 2011-present

Conferences organization committee

The First Wuxi International Statistics Forum, July 17-19, 2011, Wuxi, China.

GaTech/Emory Bioinformatics Conference. November 10-12, 2011, Atlanta, GA.

Reviewer for journals:

American Journal of Human Genetics, Annals of Applied Statistics, Annals of Human Genetics, Bayesian Analysis, Bioinformatics, Biostatistics, BMC Bioinformatics, Biometrics, Biometrika, BioTechniques, Cancer Epidemiology, Biomarkers and Prevention, Cancer Informatics, Cancer Research, Circulation, Computational Statistics, Diabetes, Frontiers in Biosciences, Genetics, Genetic Epidemiology, Genome Biology, Genomics, Human Genetics, Human Molecular Genetics, Human Heredity, IEEE IEEE/ACM Transactions on

Computational Biology and Bioinformatics, The International Journal of Bioinformatics Research and Applications, Journal of the American Statistical Association, Journal of Bioinformatics and Computational Biology, Journal of Clinical Oncology, Journal of Computational and Graphical Statistics, Mammalian Genome, Nucleic Acids Research, OMICS, Pharmacogenomics, PLoS Computational Biology, PLoS Genetics, PLoS ONE, PLoS Physiology, Statistical Application in Genetics and Molecular Biology, Scandinavian Journal of Statistics, Statistica Sinica.

Reviewer for conference:

RECOMB 2007, RECOMB-Regulatory Genomics 2007, The International Conference on Bioinformatics (InCoB) 2007, Life System Modeling and Simulation (LSMS) 2007, Bioinformatics Research and Development (BIRD) 2008, ISMB/ECCB 2009, ISMB 2010.

Reviewer for book proposal:

Springer.

Grant review:

Biostatistics reviewer, NIH study section CICS, 2005-2011.
Biostatistics reviewer, NIH study section HEMT-D, 2008.
Reviewer, NIH study section VH-E 90S, 2009.
Reviewer NIH study section ZRG1 HDM-S (02) 2011.
Mail reviewer, NIH RC1 and RC4 grants 2009, 2010.
Reviewer, Wellcome Trust Research Career Development Fellowship 2004-2005.
Reviewer, NSF 2005, 2006.
Reviewer, University of Michigan CCMB Pilot Grant. 2006 - 2008.

Teaching

At University of Michigan

BIOSTAT 601 Probability and Distribution Theory.
Fall 2006 (enrollment: 57 (2 sessions)),
Fall 2008 (enrollment: 41),
Fall 2009 (enrollment: 40).

BIOSTAT 646 / STAT 545 / BIOINFO 545 Data Analysis in Molecular Biology (inaugural class designed by myself):

Winter 2004 (enrollment: 25),
Winter 2005 (enrollment: 25),
Winter 2008 (enrollment: 14),
Winter 2009 (enrollment: 18).

BIOSTAT 666 Statistical Models and Numerical Methods in Human Genetics.
Winter 2008 (enrollment: 14).

BIOSTAT 682 Applied Bayesian Inference.

Winter 2005 (enrollment: 16),
Winter 2006 (enrollment: 27).

BIOSTAT 830 Special Topics in Statistical Genetics.
(inaugural class designed by myself):
Winter 2006 (enrollment: 9).

Postdoctoral fellows supervised

Li Li (2011- present)
Meng Zhao (2011 - present)
Rendong Yang (2011 - present)

Doctoral committee service

At Emory University

Committee member:
Hesen Peng

At University of Michigan

Committee chair or co-chair:

Jihao Zhou, Biostatistics (co-chair with Lang Li 2007)
Hyung Won Choi, Biostatistics (co-chair with Debashis Ghosh 2008)
Ming Hu, Biostatistics (2010)
Shyam Gopalakrishnan, Biostatistics (2011)

Committee member:

Lei Xu, Biostatistics (2007)
Xiaoxi Zhang, Biostatistics (2007)
Dongxiao Zhu, Bioinformatics (2006)
Ji Chen, Bioinformatics (2006)
Yili Chen, Bioinformatics (2007)
Xing Li, Bioinformatics (2008)
Shankar Subramanian, Bioinformatics (2008)
Ben-Yang Liao, EEB (2008)
Hokeun Sun, Statistics (2008)
Victorya Strumba, Bioinformatics (2009)
Jinyao Zhang, Physics (2009)
Jin Zheng, Biostatistics (2009)
Zhi Wang, Ecology and Evolutionary Biology (2010)
Grace Lin, Cellular and Molecular Biology

Bioinformatics doctoral preliminary examination committee service

Jung Hei Kim (2004)
Andrew Hodges (2006)
Jenna Vanliere (2007)

Sirarat Sarntivijai (2007)
Gang Su (2008)
Junguk Hur (2008)
Youngsheng Huang (2008)
Yu-Hsuan Lin (2008)
Ryan Welch (2008)
Arun Manoharan (2008)
Yasin Senbabaoglu (2008)
Erin R. Shellman (2009)
Chunchao Zhang (2009)
Lee Sam (2009)
Yan Zhang (2009)