Wei Sun

Last updates: Nov 16, 2014

University of North Carolina-Chapel Hill
Department of Biostatistics
4115E McGavran-Greenberg Hall
Department of Genetics
5043 Genetic Medicine Building
Chapel Hill, North Carolina 27599-7420

919-966-7266
weisun@email.unc.edu
http://www.bios.unc.edu/~weisun

Education

Ph.D.	University of California at Los Angeles Advisor: Dr. Ker-Chau Li	Statistics	2007
M.S.	University of California at Los Angeles	Statistics	2004
B.S.	Peking University, Beijing, China	Statistics	2002

Professional Experience

Department of Biostatistics, Department of Genetics, Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill

Assistant Professor	July 2007 – May 2013
Associate Professor	June 2013 – present

UNC Center for Environmental Health and Susceptibility
Director of Biostatistics and Bioinformatics Facility Core
July 2013 - present

Honors

Junior Faculty Development Awards, UNC-Chapel Hill, 2010
University Research Council Award, UNC-Chapel Hill, 2008
University of California Fellowship, UCLA, 2002-2007

Professional Service

Associate Editor:	JASA, Application and Case Studies, 2012-
JSM Program Chair:	Biometrics Section, JSM, 2013
Organizer for:	Invited Session for JSM 2011 Invited Session for ENAR 2011-2014
Publication Referee:	Annual of Statistics / Annual of Applied Statistics / Bioinformatics / Biometrics / Biostatistics / BMC Bioinformatics / BMC Genomics /

Genetic Epidemiology / Genetics / Genetics Research / International Journal of Biostatistics / JASA / Nucleic Acid Research (NAR) / Nature Methods / PLoS ONE / PNAS/ SAGMB / Society for Industrial and Applied Mathematics Book Series / Statistics and Its Interface / Statistics in Medicine / Statistics Surveys / Transactions on Computational Biology and Bioinformatics (IEEE)

Grant Reviewer: Mail reviewer for NIH RC4 applications - Recovery Act Limited Competition

Ad hoc reviewer for NIH Epidemiology of Cancer [EPIC] study section, June 2013

Ad hoc reviewer for NCI Omnibus "Biomarkers" ZCA1 SRLB-5 (J3), December 2013

Memberships of Professional Societies

2007-present: American Statistical Association
2007-present: American Society of Human Genetics
2008-present: International Biometric Society
2008-present: International Chinese Statistical Association

Research Interest

Statistical genetics, genomics, and computational biology

More generally, how life works. More specifically, gene expression quantitative trait loci (eQTL), RNA-seq, ChIP-seq, copy number variation, dimension reduction, variable selection, Bayesian Network, personalized genomic medicine, and genetic/epigenetic basis of Cancer.

Bibliography

Book Chapters

Wei Sun and Yijuan Hu (2014),
Mapping of Expression Quantitative Trait Loci Using RNA-seq Data
Statistical Analysis of Next Generation Sequencing Data
Springer Series: Frontiers in Probability and the Statistical Sciences
Datta, Somnath, Nettleton, Dan (Eds.)

Wei Sun and Min Jin Ha (2015)
eQTL and Directed Graphical Model
Integrating omics data: statistical and computational methods
Springer
George Tseng, Debashis Ghosh and Xianghong Jasmine Zhou (Eds.)

Papers under revision

1. **Sun, W.**, Liu, Y., Crowley, J. J., Chen, T. H., Zhou, H., Chu, H., ... & de Villena, F. P. M. (2014). IsoDOT Detects Differential RNA-isoform Usage with respect to a Categorical or Continuous Covariate with High Sensitivity and Specificity. arXiv preprint arXiv:1402.0136.

2. Ha MJ, **Sun W**, & Xie, J (2014). PenPC: A Two-step Approach to Estimate the Skeletons of High Dimensional Directed Acyclic Graphs. arXiv preprint arXiv:1405.1603

3. Chen TH, **Sun W**, and Fine JP (2014). Designing penalties for high dimensional regression,

4. Hu YJ, **Sun W**, Tzeng JY, and Perou CM (2014) Proper Use of Allele-Specific Expression Improves Statistical Power for *cis*-eQTL Mapping with RNA-Seq Data

5. Rashid, NU, **Sun, W**, and Ibrahim, JG (2014). A Statistical Model to Assess (Allele-Specific) Associations Between Gene Expression and Epigenetic Marks Using Sequencing Data

6. Chen TH and **Sun W** (2014). Prediction Of Cancer Drug Sensitivity Using High-Dimensional Genomic Features

Peer Reviewed Publications (Statistical Methods)

1. Wilkerson, M. D., Cabanski, C. R., **Sun, W.**, Hoadley, K. A., Walter, V., Mose, L. E., ... & Hayes, D. N. (2014). Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. Nucleic acids research, 42(13), e107-e107.

2. Ha, MJ, and **Sun, W** (2014)
   Partial correlation matrix estimation using ridge penalty followed by thresholding and reestimation
   Biometrics, in press

3. Zou F, **Sun W**, Crowley JJ, Zhabotynsky V, Sullivan PF, Pardo-Manuel de Villena FF (2014)
   A Novel Statistical Approach for Jointly Analyzing RNA-seq Data from F1 Reciprocal Crosses and Inbred Lines
   Genetics, 197(1), 389-399

4. Hu, YJ, Lin, DY, **Sun, W**, and Zeng, D (2014)
   A Likelihood-Based Framework for Association Analysis of Allele-Specific Copy Numbers
   Journal of the American Statistical Association, in press,

5. Rashid, NU, **Sun, W**, and Ibrahim, JG (2014).
   Some Statistical Strategies for DAE-seq Data Analysis: Variable Selection and Modeling Dependencies among Observations.

Journal of the American Statistical Association, 109(505):78-94,

6. Szatkiewicz JP, Wang W, Sullivan PF, Wang W, **Sun W** (2013)
   Improving detection of copy-number variation by simultaneous bias correction and read-depth segmentation.
   Nucleic Acids Res. 1;41(3):1519-32. PMCID: PMC3561969

7. **Sun W**, Hu Y (2013)
   eQTL Mapping Using RNA-seq Data.
   Statistics in Bioscience, in 2013 May1;5(1):198-219. PMCID: PMC3650863

8. Lee W, Du Y, **Sun W**, Hayes D.N., Liu Y (2012)
   Multiple Response Regression for Gaussian Mixture Models with Known Labels.
   Statistical Analysis and Data Mining, 5(6), 493-508.

9. Zhang X, Huang S, **Sun W**, Wang W (2012)
   Rapid and robust resampling-based multiple-testing correction with application in a genome-wide expression quantitative trait loci study.
   Genetics. 2012 Apr;190(4):1511-20. PMCID: PMC3316660.

10. **Sun W** (2012)
    A statistical framework for eQTL mapping using RNA-seq data.
    Biometrics. 2012 Mar;68(1):1-11. Epub 2011 Aug 12. PMCID: PMC3218220.

11. **Sun W**, Li L (2012)
    Multiple loci mapping via model-free variable selection.
    Biometrics. 2012 Mar;68(1):12-22. Epub 2011 Aug 12. PMCID: PMC3218235.

12. Rashid NU, Giresi PG, Ibrahim JG, **Sun W***, Lieb JD* (2011)
    ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions.
    Genome Biology, 12(7):R67 (20 pages), PMCID: PMC3218829.
    * co-corresponding author

13. Yu T, Peng H, **Sun W** (2011)
    Incorporating Nonlinear Relationships in Microarray Missing Value Imputation.
    IEEE/ACM Trans Comput Biol Bioinform. 2011 May-Jun;8(3):723-31.

14. Chu H, Nie L, Chen Y, Huang Y, **Sun W** (2010)
    Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: Methods for the absolute risk difference and relative risk.
    Stat Methods Med Res. 2010 Dec 21 (13 pages).

15. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, **Sun W**, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Børresen-Dale AL, Kristensen VN (2010)
    Allele-specific copy number analysis of tumors.
    Proc Natl Acad Sci U S A. 2010 Sep 28;107(39):16910-5. PMCID: PMC2947907.

16. **Sun W**, Ibrahim JG, Zou F (2010)

Genome-wide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression.
Genetics. 2010 May;185(1):349-59. Epub 2010 Feb 15. PMCID: PMC2870969.

17. Shen Y, **Sun W**, Li KC (2010)
Dynamically weighted clustering with noise set.
Bioinformatics. 2010 Feb 1;26(3):341-7. Epub 2009 Dec 9. PMCID: PMC2815660.

18. **Sun W** and Wright FA (2010)
A geometric interpretation of the permutation p-value and its application in eQTL studies.
Annals of Applied Statistics, 4(2), 1014-1033

19. **Sun W**, Wright FA, Tang Z, Nordgard SH, Van Loo P, Yu T, Kristensen VN, Perou CM (2009)
Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res*. 2009 Sep;37(16):5365-77. Epub 2009 Jul 6. PMCID: PMC2935461.

20. **Sun W**, Buck MJ, Patel M, Davis IJ (2009)
Improved ChIP-chip analysis by a mixture model approach.
BMC Bioinformatics. 2009 Jun 7;10:173 (13 pages). PMCID: PMC2700807

21. **Sun W***, Xie W*, Xu F, Grunstein M, Li KC (2009)
Dissecting nucleosome free regions by a segmental semi-Markov model.
PLoS One. 2009;4(3):e4721 (10 pages). Epub 2009 Mar 6. PMCID: PMC2648986.
* Co-first authors

22. Wu T*, **Sun W***, Yuan S, Chen CH, Li KC (2008)
A method for analyzing censored survival phenotype with gene expression data.
BMC Bioinformatics. 2008 Oct 6;9:417 (11 pages). PMCID: PMC2579309.
* Co-first authors

23. **Sun W***, Yuan S*, Li KC (2008)
Trait-trait dynamic interaction: 2D-trait eQTL mapping for genetic variation study.
BMC Genomics. 2008 May 23;9:242 (13 pages). PMCID: PMC2432080.
* Co-first authors

24. **Sun W**, Yu T, Li KC (2007)
Detection of eQTL modules mediated by activity levels of transcription factors. *Bioinformatics*. 2007 Sep 1;23(17):2290-7.

25. Yu T, Ye H, **Sun W**, Li KC, Chen Z, Jacobs S, Bailey DK, Wong DT, Zhou X (2007)
A forward-backward fragment assembling algorithm for the identification of genomic amplification and deletion breakpoints using high-density single nucleotide polymorphism (SNP) array.
BMC Bioinformatics. 2007 May 3;8:145 (11 pages). PMCID: PMC1868765.

26. Yu T, **Sun W**, Yuan S, Li KC (2005)
Study of coordinative gene expression at the biological process level.

Bioinformatics, 2005 Sep 15;21(18):3651-7.

27. Li KC, Liu CT, **Sun W**, Yuan S, Yu T (2004)
    A system for enhancing genome-wide coexpression dynamics study.
    Proc Natl Acad Sci U S A. 2004 Nov 2;101(44):15561-6. PMCID: PMC524832.

Peer Reviewed Publications (Applied-Collaborative)

28. James J Crowley*, Vasyl Zhabotynsky*, **Wei Sun***, ..., Fernando Pardo-Manuel de Villena (2014)
    Pervasive Allelic Imbalance Revealed By Allele-Specific Gene Expression In Highly Divergent Mouse Crosses.
    Nature Genetics, in press
    * Co-first authors

29. Fred Wright, Patrick Sullivan, Andrew Brooks, Fei Zou, **Wei Sun** et al. (2014)
    Heritability and genomics of gene expression in peripheral blood,
    Nature Genetics, 46(5), 430-437.

30. Holley DW, Groh BS, Wozniak G, Donohoe DR, **Sun W**, Godfrey V, Bultman SJ. (2014)
    The BRG1 chromatin remodeler regulates widespread changes in gene expression and cell proliferation during B cell activation.
    J Cell Physiol. 2014 Jan;229(1):44-52.

31. Moeller BC, Recio L, Green A, **Sun W**, Wright FA, Bodnar WM, Swenberg JA (2013)
    Biomarkers of Exposure and Effect in Human Lymphoblastoid TK6 Cells Following [13C2]-Acetaldehyde Exposure.
    Toxicological Sciences, 2013 May;133(1):1-12 PMCID: PMC3627555

32. Calabrese JM, **Sun W**, Song L, Mugford J, Williams L, Yee D, Starmer J, Mieczkowski P, Crawford G, Magnuson T (2012)
    Site-specific silencing of regulatory elements as a mechanism of X-inactivation.
    Cell, 2012 Nov 21;151(5):951-63. PMCID: PMC3511858

33. Donohoe DR, Collins LB, Wali A, Bigler R, **Sun W**, Bultman SJ (2012)
    The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation.
    Mol Cell. 2012 Nov 30;48(4):612-26. PMCID: PMC3513569

34. Zhao N, Ang MK, Yin XY, Patel MR, Fritchie K, Thorne L, Muldrew KL, Hayward MC, **Sun W**, Wilkerson MD, Chera BS, Hackman T, Zanation AM, Grilley-Olson JE, Couch ME, Shockley WW, Weissler MC, Shores CG, Funkhouser WK, Olshan AF, Hayes DN (2012)
    Different cellular p16(INK4a) localisation may signal different survival outcomes in head and neck cancer.
    Br J Cancer, 2012 Jul 24;107(3):482-90. PMCID: PMC3405208

35. **Collaborative Cross Consortium** (2012)
    The genome architecture of the Collaborative Cross mouse genetic reference population.

Genetics. 190(2):389-401. PMCID: PMC3276630

36. Xia K, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, Zou F, **Sun W**, Sullivan PF, Wright FA (2012)
SeeQTL: a searchable database for human eQTLs.
Bioinformatics. 2012 Feb 1;28(3):451-2. PMCID: PMC3268245.

37. **Sun W**, Lee S, Zhabotynsky V, Zou F, Wright FA, Crowley JJ, Yun Z, Buus RJ, Miller DR, Wang J, McMillan L, Pardo-Manuel de Villena F, Sullivan PF (2012)
Transcriptome atlases of mouse brain reveals differential expression across brain regions and genetic backgrounds.
G3 (Bethesda). 2012 Feb;2(2):203-11. PMCID: PMC3284328

38. Donohoe DR, Garge N, Zhang X, **Sun W**, O'Connell TM, Bunger MK, Bultman SJ (2011)
The Microbiome and Butyrate Regulate Metabolism and Autophagy in the Mammalian Colon.
Cell Metabolism, Vol. 13, 517-526. PMCID: PMC3099420

39. Jones MD, **Sun W**, and Aitken MD (2011)
Multiple DNA extractions coupled to stable-isotope probing of anthracene-degrading bacteria in contaminated soil.
Applied and Environmental Microbiology 2011, Vol. 77, 2984-2991.

40. Campos M*, **Sun W***, Yu F*, Barbalic M, Tang W, Chambless LE, Wu KK, Ballantyne C, Folsom AR, Boerwinkle E, Dong JF (2011)
Genetic Determinants of Plasma von Willebrand Factor Antigen Levels: A Target Gene SNP and Haplotype Analysis of ARIC Cohort.
Blood 2011, Vol. 117, 5224-5230. PMCID: PMC3109544
* Co-first authors

41. Dehghan A, Dupuis J, …, **Sun W**., …, Chasman DI (2011)
Meta-analysis of genome-wide association studies in over 80,000 subjects identifies multiple loci for C-reactive protein levels.
Circulation 2011, Vol. 123, 731-738

42. Wright FA, Strug LJ, Doshi VK, ..., **Sun W**, ..., Knowles MR, Cutting GR (2011)
Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2.
Nat Genet 2011 Jun;43(6):539-46. PMCID: PMC3296486.

43. Gatti DM, Lu L, Williams RW, **Sun W**, Wright FA, Threadgill DW, Rusyn I (2011)
MicroRNA Expression in the Livers of Inbred Mice.
Mutation Research, 2011 Sep 1;714(1-2):126-33. Epub 2011 May 14

Software Development

http://www.bios.unc.edu/~weisun/software.htm

- asSeq
  Allele-specific eQTL mapping

- isoform
  A set of tools for RNA isoform study using RNA-seq data

- BPrimm
  Bayesian and Penalized regression in multiple loci mapping. It includes a set of tools for simultaneously multiple loci mapping, and two novel methods named the Bayesian adaptive Lasso and the Iterative Adaptive Lasso

- genoCN
  Simultaneously dissect copy number states and genotypes using the data from high density SNP arrays

- mixer
  A mixture model approach to analyze ChIP-chip or ChIP-seq data, also with some utility functions to process DNA sequence data.

- permuteP
  Exact permutation p-value calculation for case-control study of collapsing a group of rare SNPs, and approximate permutation p-value calculation for QTL mapping in experimental cross

- eMap
  eQTL computation, visualization, eQTL module, integrated studies of complex trait, gene expression and genetic markers, etc.

- censorSIR
  Tools for applying Slice Inverse Regression in censored data.

Invited Talks

1. Computational Biology and Statistics Workshop, in celebrating the 10th anniversary of UCLA Department of Statistics, *October 19-20, 2008*, Title: Genome-wide Multiple Loci Mapping in Experimental Crosses of Inbred Strains Using Dense Genetic Markers
2. Department of Biostatistics, Emory University, *September 3, 2009*, Title: Integrated Study of Copy Number States and Genotype Calls Using High Density SNP Arrays
3. Department of Biostatistics, University of Washington St Louis, *February 19, 2010*, Title: Integrated Dissection of Copy Number States and Genotype Calls

4. Workshop on Statistical Frontiers, Institute of statistical science, Academia Sinica, Taipei, Taiwan, *December 2010*, Title: Statistical Methods for eQTL Mapping using RNA-seq Data
5. 2010 Annual Meeting of Chinese Statistical Society and International Statistical Conference, National Central University, Taiwan, *December 2010*, Title: Statistical Strategies for Some Variable Selection Problems in Genetic Studies
6. Department of Biostatistics, University of Texas at Houston, *April 4, 2011*, Title: Allele-specific eQTL Mapping, Title: Allele-specific eQTL Mapping
7. The Jackson Laboratory, *May 24, 2011*, Title: Allele-specific expression by RNA-seq
8. International Chinese Statistical Association 2011 Applied Statistics Symposium, *June 27, 2011*, New York City, Title: Statistical Methods for eQTL Mapping using RNA-seq Data
9. Department of Statistics, Duke University, *October 14, 2011*, Title: Multiple loci mapping by penalized regression
10. Department of Statistics, UCLA, *Nov 1st, 2011*, Title: Multiple loci mapping by penalized regression
11. Department of Biostatistics, University of Pittsburgh, *Oct 11, 2012*, Title: Statistical methods for RNA seq studies
12. Department of Statistics, University of Illinois at Urbana-Champaign, *Oct 18, 2012*, Title: Statistical methods for RNA seq studies
13. NIEHS, *Nov 20, 2012*, Title: Statistical methods for RNA seq studies
14. NC State University, Biostatistics Working Group, *Feb 14, 2013*, Title: PenPC: A Two-step Approach to Estimate the Skeletons of High Dimensional Directed Acyclic Graphs
15. Peking University, School of Mathematical Sciences, Young Mathematician Forum in celebrating of 100th anniversary of the School of Mathematical Sciences (SMS) at Peking University, *June 18-21, 2013*, Title: Statistical Methods for Cancer Genomics
16. SAMSI, LDHD Transition Workshop, *May 05 2014*, Title: Estimation of High Dimensional Directed Acyclic Graphs using eQTL Data
17. Department of Biostatistics, Duke University, *May 15 2014*, Title: Statistical Methods for Cancer Genomics
18. Statistical Society of Canada 2014 Annual Meeting, *May 27 2014*, Toronto, Canada, Title: A study of RNA-seq data in Cancer Patients
19. Department of Applied and Computational Mathematics and Statistics, *Oct 27 2014*, University of Notre Dame, Title: Statistical Methods for Cancer Genomics
20. Triangle Statistical Genetics Conference, SAS Campus, *Oct 31 2014*, Title: Statistical methods to exploit allele-specific and isoform-specific information from RNA seq data

Teaching

Graduate Course

Bios784 Introduction to computational biology, 2009 spring, 2011 spring, 2013 spring
Bios735 Statistical Computing, 2013 fall
Bios663 Intermediate linear models, 2010 spring, 2012 spring, 2014 spring

Direction of Doctoral Research

John C. Schwarz (2006-2010), "Enhancing eQTL Analysis Techniques with Special Attention to the Transcript Dependency Structure", co-advised with Dr. Fred Wright

Naim Rashid (2007-2013), "Model-based approaches for the detection of biologically active genomic regions from next generation sequencing data", co-advised with Dr. Joe Ibrahim

Min-Jin Ha (2008-2013), "Estimation of directed and non-directed gene expression network using gene expression and genetic data"

Ting-Huei Chen (2009-2014), "Penalized Estimation Methods and Their Applications in Genomics and Beyond", co-advised with Dr. Jason Fine

Matt Psioda (2013-), "Mixture cell type deconvolution", co-advised with Dr. Joe Ibrahim

Jenny Yang (2014-), "Model-free estimation of graphic model", co-advised with Dr. Donglin Zeng

Doug Wilson (2014-), co-advised with Dr. Joe Ibrahim

Direction of Master's Research

Zhengzheng Tang (2009), Integrated study of copy number states and genotype calls using high-density SNP arrays

This work won Department of Biostatistics Regina Elandt-Johnson Award - Best Masters paper completed in 2009

Guanhua Chen (2010), Statistical methods for analyzing customized copy number variation array

Ni Zhao (2012), eQTL Mapping Using RNA-seq Data

Direction of BSPH (Bachelor of Science in Public Health) Honors Paper

Mary Cooter (2011), A Two-step Approach for Accurate Detection of Copy Number Variations

PhD Dissertation Committee

	Student	Department	Advisor	Graduation
1	Ramon I. Garcia	Biostatistics	Joe Ibrahim & Hongtu Zhu	2009
2	Daniel Gatti	Environmental Sciences	Ivan Rusyn	2010 March
3	Seo Young Park	Statistics	Yufeng Liu	2010 April
4	Paul Giresi	Biology	Jason Lieb	2010 May
5	YuYing Xie	Genetics	David Threadgill	2010 July
6	Von Walter	Biostatistics	Fred Wright	2010 July
7	Yijuan Hu	Biostatistics	Danyu Lin	2011 May
8	Yi Gong	Biostatistics	Fei Zou	2011 June
9	Yihui Zhou	Biostatistics	Fred Wright	2011 Nov.
10	Eric Lock	Statistics	Andrew Nobel & Steve Marron	2012 April
11	Colin Lickwar	Biology	Jason Lieb	2012 July
12	Thomas Clarke	Biology	Vison Todd	2012 Oct
13	Ja-an Lin	Biostatistics	Hongtu Zhu & Joe Ibrahim	2013 July
14	Lan Liu	Biostatistics	Michael Hudgens	2013 July
15	Khondker Zakaria	Biostatistics	Hongtu Zhu & Joe Ibrahim	2013 Aug
16	Wonil Chung	Biostatistics	Fei Zou	2013 Nov
17	Gene Urrutia	Biostatistics	Michael Wu	2013 Nov
18	Andrea Byrnes	Biostatistics	Yun Li	2013 Nov
19	Ni Zhao	Biostatistics	Michael Wu	2013 Dec
20	Zhengzheng Tang	Biostatistics	Danyu Lin	2014 June
21	Guanhua Chen	Biostatistics	Michael Kosorok	2014 June
22	Weibo Wang	Computer Science	Wei Wang	2014 Dec
23	Shunping Huang	Computer Science	Wei Wang	2014 Dec
24	Matthew Weiser	Genetics	Terry Furey	In progress
25	Guosheng Zhang	Genetics	Yun Li	In progress
25	Yuying Xie	Genetics	William Valdar and Yufeng Liu	In progress
26	Zhaoyu Yin	Biostatistics	Fei Zou	2014 Nov
27	Greg Keele	BCB (Bioinformatics and Computational biology)	William Valdar	In progress
28	Dan Oreper	BCB	William Valdar	In progress
28	Wei Cheng	Computer Science	Wei Wang	In progress
29	Chen-Ping Fu	Computer Science	Leonard McMilan	In progress
30	Pratyaydipta Rudra	Biostatistics	Fred Wright and Andrew Nobel	In progress

Research Grant

Pending

1 R01 CA197485-01 (Sun) 7/1/15 6/30/19
National Institutes of Health
Some Computational Tools for Personalized Cancer Therapeutics
We propos to develop computational methods to identify new cancer prognosis biomarkers based on allele-specific gene expression, to predict cancer drug sensitivity while accounting for intra-tumor heterogeneity, and to prioritize cancer drug targets using graphical models.
Role: Principal Investigator

Ongoing Research Support

1 R01 GM105785-01 (Sun) 5/15/2014 4/30/18
National Institute of General Medical Sciences
Statistical Methods for RNA-seq data analysis
We propose to develop statistical methods for joint study of germline DNA polymorphisms and allele-specific expression (ASE) obtained from RNA-seq data. Since germline DNA polymorphisms are stable across tissues and developmental stages, inclusion of DNA information will help us establish more reliable biomarkers for patients' clinical care. More specifically, we will study the genetic basis of ASE in both normal and tumor tissues, dissect genetic and parent-of-origin effects on ASE in human cell lines, and identify genes that escape X inactivation in both mouse reciprocal cross and human cell lines.
Role: Principal Investigator

5 R03 CA167684-02 (Sun) 5/1/2012 4/30/14
National Cancer Institute
Dissect Genetic Basis of Cancer Using Allele-Specific Gene Expression
In this project, we will apply our methods to analyze a genomic dataset of 248 colorectal cancer patients derived from The Cancer Genome Atlas project. Our methods and the results of our real data analysis will benefit the detection, diagnosis, treatment, and prognosis of colorectal cancer, as well as other types of cancers.
Role: Principal Investigator

5 P30 ES010126-13 (Swenberg) 4/1/2010 3/31/15
National Institute of Environmental Health Sciences
UNC-CH Center for Environmental Health & Susceptibility - Facility Core 2: Biostatistics & Bioinformatics
The UNC-CH Center on Environmental Health and Susceptibility brings population science, medical and biomedical researchers together to examine major issues in environmental health resulting from gene-environment interactions that affect an individual's susceptibility to disease.
Role: Core Director

5 R01 GM074175-07 (Zou) 4/1/06 8/31/15
National Institute of General Medical Sciences
Robust Methods for Complex Trait Mapping with Collaborative Cross
The ultimate goal of the proposal is to provide scientists working on collaborative cross (CC) mice with a statistical analysis platform, which contains specially designed analytical

tools for CC mouse data, ranging from simple univariate analysis to more complicated multivariate and longitudinal data analysis, and highly complex integrated high-dimensional data analysis.
Role: Co-Investigator

2 R01 CA082659-14A1 (Lin) 4/1/2008 1/31/17
National Cancer Institute
Statistical Methods in Chronic Disease Research
The broad, long-term objectives of this research are the developments of innovative and high-impact statistical methods for the design and analysis of chronic disease studies, with an emphasis on genomics.
Role: Co-Investigator

5 U24 CA143848-05 (Perou) 9/29/2009 7/31/14
National Cancer Institute
Gene Expression Patterns in Human Tumors Identified using Transcript Sequencing
Our study of genome-wide transcript regulation with chromatin organization will provide a critical portrait of the cancer genome that can be integrated with (and indeed can sometimes generate) other important data, including mutations and copy number events.
Role: Biostatistician

5 R01 CA149569-05 (Liu) 2/1/2010 12/31/14
National Cancer Institute
Flexible Statistical Machine Learning Techniques for Cancer-Related Data
The goal of this project is to develop a host of new statistical learning techniques for solving complicated learning problems. The proposed techniques will be applied to cancer research data analysis.
Role: Co-Investigator

1 R01 HG006292-03 (Li) 8/23/2011 5/31/16
National Human Genome Research Institute
Design and Analysis of Sequencing-based Studies for Complex Human Traits
We will establish a comprehensive statistical framework for the design and analysis of sequencing-based studies.
Role: Co-Investigator

5 R01 GM070335-15 (Ibrahim) 9/15/2011 8/31/15
National Institute of General Medical Sciences
Bayesian Approaches to Model Selection for Survival Data
In this proposal, we develop Bayesian methodology for high dimensional genomic data. The overarching theme in this proposal is that we develop several novel statistical methods for motif discovery in genomic sequence data.
Role: Co-Investigator

1 R01 MH101819-01 (Wright) 8/15/2013 7/31/16
North Carolina State University
Systems Approaches to Link Tissue-Specific Expression to Disease
This application aims to provide the tools, analysis, and framework to identify which DNA variants affect the expression of genes in a manner that is common to many tissues, and

which are specific to one or a few tissues. The results will be used to gain a better understanding of the complex patterns of DNA variation in causing complex disease.
Role: Co-Investigator

Completed Research Support

N 01 HC-55015-45-0-1 (Chambless) 12/1/2005 1/31/12
National Heart, Lung and Blood Institute
Atherosclerosis Risk in Communities (ARIC)
The ARIC Study is a prospective study conducted in four US communities to investigate the etiology and natural history of atherosclerosis disease and measure variation in cardiovascular risk factors. ARIC is designed to investigate the etiology and natural history of atherosclerosis, the etiology of clinical atherosclerotic diseases, and variation in cardiovascular risk factors, medical care and disease by race, sex, place, and time (15,800 total patient participants).
Role: Biostatistician

5 U01 HG004402-02 (Boerwinkle) 7/6/2007 5/31/11
National Human Genome Research Institute
Genome-wide Association for Gene-Environment Interaction Effects Influencing CHD
For this proposed research, we will leverage the full scope of the ARIC study resources, including the currently-funded genome-wide association ancillary study to identify gene by environment interactions affecting incident CHD and other heart, lung and blood -related phenotypes.
Role: Biostatistician

RD-83382501 (Rusyn) 4/1/2008 3/31/13
Environmental Protection Agency
Carolina Center for Computational Toxicology - Project 2
The objective of this proposal is to create The Carolina Center for Computational Toxicology. We present a clear plan for an effective, broad and interdisciplinary effort to devise novel tools, methods and knowledge that will utilize ToxCast and other publicly available data to assist the EPA in achieving the goals of protecting the environment and human health.
Role: Co-Investigator

5 R01 HL095396-04 (Knowles) 9/24/2008 7/31/12
National Institute of Heart, Lung, and Blood
Molecular Phenotypes for Cystic Fibrosis Lung Disease
Cystic fibrosis (CF) is a recessive genetic disorder caused by mutations in CFTR. This project holds great promise for defining a robust molecular phenotype for CF lung disease, which relates to prognosis, and new targets for therapy.
Role: Co-Investigator

GIL200811.0020 (Lin) 7/1/2009 11/30/11
UNC Gillings Innovation Laboratories
Gillings Innovative Laboratory in Statistical Genomics
The Gillings Innovation Laboratory in Statistical Genomics will develop novel and high-impact statistical methods for the design and analysis of genetic association studies, so as to maintain and enhance UNC's leading status in several areas of epidemiology and genetics.
Role: Co-Investigator

5 RC2 MH089951-02 (Sullivan) 9/30/2009 8/31/12
National Institute of Mental Health
Integration of Genomics and Transcriptomics in Normal Twins and Major Depression
Our goals are to develop a comprehensive understanding of the genomics of transcription in normalcy and then to discover DNA and RNA biomarkers for major depressive disorder.
Role: Co-Investigator

5 P50 MH090338-02 (Pardo-Manuel de Villena) 9/30/2009 8/31/11
National Institute of Mental Health
An Interdisciplinary Program for Systems Genomics of Complex Behaviors
We propose a highly ambitious yet realistically attainable goal: to align existing expertise at UNC-CH into a center of excellence in order to develop as a resource and demonstrate the utility of the murine Collaborative Cross (CC) to delineate genetic and environmental determinants of complex phenotypes drawn from psychiatry, the most intractable set of problems in all of biomedicine.
Role: Co-Investigator