

Wei Sun

Last updates: September 15, 2021

Fred Hutchinson Cancer Research Center
Biostatistics Program
1100 Fairview Ave N, M2-B845
P.O. Box 19024
Seattle, WA 98109-1024

206-667-3188
wsun@fredhutch.org

Education

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

Professional Experience

Public Health Division, Fred Hutchison Cancer Research Center, Seattle, WA
Associate Professor September 2015 – December 2018
Professor January 2019 – Present

Department of Biostatistics, University of North Carolina, Chapel Hill
Affiliated Associate Professor September 2016 – August 2019
Affiliated Professor September 2019 – Present

Department of Biostatistics, University of Washington
Affiliated Associate Professor April 2017 – Present

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 – August 2016

UNC Center for Environmental Health and Susceptibility
Director of Biostatistics and Bioinformatics Facility Core
July 2013 - August 2015

Honors

Elected Fellow, American Statistical Association, 2017
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-

ENAR 2016 Program Chair, ENAR Regional Advisory Board for 2016-2018

JSM Program Chair: Biometrics Section, JSM, 2013

Organizer for: Invited Session for JSM 2011, 2016, 2018, 2019, 2021
Invited Session for ENAR 2011-2015, 2017, 2018

Publication Referee: Annual of Statistics / Annual of Applied Statistics /
Bioinformatics / Biometrics / Biostatistics /
BMC Bioinformatics / BMC Genomics /
Genetic Epidemiology / Genetics / Genetics Research /
Genome Biology / JASA / Nucleic Acid Research (NAR) /
Nature Genetics/ Nature Methods / Nature Communication /
PLOS ONE / PLOS Computational Biology / PLOS Genetics /
PNAS/ SAGMB / 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

NIH Epidemiology of Cancer [EPIC] study section, June 2013

NCI Omnibus "Biomarkers" ZCA1 SRLB-5, December 2013

NIH Special Emphasis Panel, March 2015

NCI Special Emphasis Panel, April 2016

NIH Special Emphasis Panel, March 2017

NCATS Special Emphasis Panel, Feb 2019

NCATS Special Emphasis Panel, Feb 2020

NIGMS MIRA grant review special panel, March 2021

External Reviewer:

NIEHS, Biostatistics and Computational Biology Branch, 2021

Memberships of Professional Societies

2007-present: American Statistical Association

2008-present: International Biometric Society

2008-present: International Chinese Statistical Association

2018-present: American Association of Cancer Research

Research Interest

Statistical genetics, genomics, and computational biology

More generally, how life works. More specifically, gene expression quantitative trait loci (eQTL), single cell or bulk tissue RNA-seq, ChIP-seq, copy number variation, cancer immunotherapy, dimension reduction, variable selection, Bayesian Network, and deep learning.

Bibliography

Book Chapters

1. **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.)
2. **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.)

Review Paper

1. Sylvia Richardson, George C. Tseng, and **Wei Sun** (2016)
Statistical Methods in Integrative Genomics,
Annual Review of Statistics and Its Application, 3:181-209.
2. Qianchuan He, Yang Liu, and **Wei Sun** (2018)
Statistical analysis of non-coding RNA data,
Cancer letters. 417, 161-167

Peer Reviewed Publications (Statistical Methods. * indicates corresponding author)

1. Liu, J., Wang, H., **Sun, W.**, & Liu, Y. (2021).
Prioritizing Autism Risk Genes using Personalized Graphical Models Estimated from Single Cell RNA-seq Data.
Journal of the American Statistical Association, in press, 1-35.
2. Molstad, A. J., **Sun, W.**, & Hsu, L. (2021).
A covariance-enhanced approach to multi-tissue joint eQTL mapping with application to transcriptome-wide association studies. *The annals of applied statistics*, 15(2), 998.
3. Jin C, Chen M, Lin DY, **Sun W** (2021)
Cell-type-aware analysis of RNA-seq data. *Nature Computational Science*, 1:253–261

4. Zhang H, Cai R, Dai J, **Sun W** (2021)
EMeth: An EM algorithm for cell type decomposition based on DNA methylation data.
Sci Rep. 11(1):5717
5. Yang J, Liu Y, Liu Y, **Sun W** (2021)
Model free estimation of graphical model using gene expression data.
The Annals of Applied Statistics. 15(1):194-207.
6. Molstad AJ, Hsu L, **Sun W** (2021)
Gaussian process regression for survival time prediction with genome-wide gene expression.
Biostatistics, 22(1) 164-180.
7. Sheikh, M. T., Ibrahim, J. G., Gelfond, J. A., **Sun, W.**, & Chen, M. H. (2021).
Joint modelling of longitudinal and survival data in the presence of competing risks with applications to prostate cancer data.
Statistical Modelling, 21(1-2), 72-94.
8. Wilson DR, Jin C, Ibrahim JG, **Sun W.** (2020)
ICeD-T Provides Accurate Estimates of Immune Cell Abundance in Tumor Samples by Allowing for Aberrant Gene Expression Patterns.
Journal of the American Statistical Association. 2020 Jul 2;115(531):1055-65.
9. Dai JY, Wang X, Wang B, **Sun W**, Jordahl KM, Kolb S, Nyame YA, Wright JL, Ostrander EA, Feng Z, Stanford JL.
DNA methylation and cis-regulation of gene expression by prostate cancer risk SNPs.
PLOS Genetics. 2020 Mar 30;16(3):e1008667.
10. Wilson DR, Ibrahim JG, **Sun W*** (2020)
Mapping Tumor-Specific Expression QTLs In Impure Tumor Samples.
Journal of the American Statistical Association, 115(529), 79-89
11. **Sun W***, Jin C, Gelfond JA, Chen MH, Ibrahim JG (2020)
Joint analysis of single-cell and bulk tissue sequencing data to infer intratumor heterogeneity.
Biometrics. 76(3):983-994.
12. Huang L, Little P, Huyghe JR, Shi Q, Harrison TA, Yothers G, George TJ, Peters U, Chan AT, Newcomb PA, **Sun W** (2020).
A Statistical Method for Association Analysis of Cell Type Compositions.
Statistics in Biosciences. 15:1-3
13. Little P, Lin DY, **Sun W*** (2019)
Associating somatic mutations to clinical outcomes: a pan-cancer study of survival time.
Genome medicine,11(1):37.
14. Zhabotynsky V, Inoue K, Magnuson T, Calabrese M, **Sun W*** (2019)

- Joint estimation of genetic and parent-of-origin effects using RNA-seq data from human.
Biometrics, 75(3), 864-74
15. Liu Y, **Sun W**, Reiner A, Kooperberg C, He Q (2019)
Genetic pathway analysis under high dimensions.
Biometrika, 106(3), 651–664
 16. Liu J, **Sun W**, Liu Y (2019)
Joint Skeleton Estimation of Multiple Directed Acyclic Graphs for Heterogeneous Population.
Biometrics, 75, 36-47
 17. Ha MJ, **Sun W** (2018)
Estimation of High-dimensional Directed Acyclic Graphs with Surrogate Intervention.
Biostatistics, doi:10.1093/biostatistics/kxy080
 18. Liu Y, He Q, **Sun W*** (2018)
Association analysis using somatic mutations.
PLoS Genetics, 14(11), e1007746.
 19. **Sun W***, Bunn P, Jin C, Little P, Zhabotynsky V, Perou CM, Hayes DN, Chen M, Lin DY (2018).
The association between copy number aberration, DNA methylation and gene expression in tumor samples.
Nucleic acids research, 46(6), pp.3009-3018.
 20. Wang W, **Sun W**, Wang W, Szatkiewicz J. (2018)
A randomized approach to speed up the analysis of large-scale read-count data in the application of CNV detection.
BMC bioinformatics. 2018 Dec;19(1):74.
 21. Liu Y, Xiong S, **Sun W**, Zou F (2017)
Joint Analysis of Strain and Parent-of-Origin Effects for Recombinant Inbred Intercrosses Generated from Multiparent Populations with the Collaborative Cross as an Example.
G3: Genes, Genomes, Genetics. g3-300483.
 22. Lee S, **Sun W**, Wright FA, Zou F (2017).
An improved and explicit surrogate variable analysis procedure by coefficient adjustment.
Biometrika, 104(2), 303-316.
 23. Silva GO, Siegel MB, Mose LE, Parker JS, **Sun W**, Perou CM, Chen M (2017).
SynthEx: a synthetic-normal-based DNA sequencing tool for copy number alteration detection and tumor heterogeneity profiling.
Genome Biology, 18(1), 66.

24. Rashid, NU, **Sun, W***, and Ibrahim, JG (2017).
A Statistical Model to Assess (Allele-Specific) Associations Between Gene Expression and Epigenetic Marks Using Sequencing Data, *Annals of Applied Statistics*, 10(4), 2254-2273.
25. Chen TH and **Sun W*** (2017).
Prediction of Cancer Drug Sensitivity Using High-Dimensional Genomic Features. *Biostatistics*, 18 (1): 1-14
26. Zhang Y, Zhou H, Zhou J, and **Sun W.** (2017)
Regression models for multivariate count data, *Journal of Computational Graphical Statistics*, 26(1), 1-13.
27. Chen TH, **Sun W***, and Fine JP (2016).
Designing penalty functions in high dimensional problems: The role of tuning parameters. *Electronic Journal of Statistics*, 10(2), 2312-2328.
28. Gelfond JA, Ibrahim JG, Chen MH, **Sun W**, Lewis K, Kinahan S, Hibbs M, Buffenstein R. (2015)
Homology cluster differential expression analysis for interspecies mRNA-Seq experiments. *Stat Appl Genet Mol Biol*. 14(6):507-16.
29. Wang, W., Wang, W., **Sun, W.**, Crowley, J. J., & Szatkiewicz, J. P. (2015).
Allele-specific copy-number discovery from whole-genome and whole-exome sequencing. *Nucleic acids research*, 43(14), e90-e90.
30. Ha MJ, **Sun W***, & Xie, J (2015).
PenPC: A Two-step Approach to Estimate the Skeletons of High Dimensional Directed Acyclic Graphs. *Biometrics*, 72(1), 146-155
31. Hu YJ, **Sun W**, Tzeng JY, and Perou CM (2015)
Proper Use of Allele-Specific Expression Improves Statistical Power for *cis*-eQTL Mapping with RNA-Seq Data, *Journal of the American Statistical Association*, 110(511) 962-974
32. **Sun W***, Liu Y, Crowley JJ, Chen TH, Zhou H, Chu H, Huang S, Kuan PF, Li Y, Miller D, Shaw G, ..., de Villena, FP (2015).
IsoDOT Detects Differential RNA-isoform Usage with respect to a Categorical or Continuous Covariate with High Sensitivity and Specificity. *Journal of the American Statistical Association*, 110(511) 975-986
33. Huang KC, **Sun W**, Wu Y, Chen M, Mohlke KL, Lange LA, Li Y. (2014)
Association studies with imputed variants using expectation-maximization likelihood-ratio tests. *PLoS One*. 10;9(11):e110679.
34. Lin JA, Zhu H, Mihye A, **Sun W**, Ibrahim JG (2014)

- Alzheimer's Neuroimaging Initiative. Functional-mixed effects models for candidate genetic mapping in imaging genetic studies.
Genet Epidemiol. 38(8):680-91
35. 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.
 36. Ha, MJ, and **Sun, W*** (2014)
Partial correlation matrix estimation using ridge penalty followed by thresholding and reestimation
Biometrics, 70(3), 762-770
 37. 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
 38. 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, 109(508):1533-1545,
 39. 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,
 40. 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. PMID: PMC3561969
 41. **Sun W***, Hu Y (2013)
eQTL Mapping Using RNA-seq Data.
Statistics in Bioscience, in 2013 May1;5(1):198-219. PMID: PMC3650863
 42. 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.
 43. 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. PMID: PMC3316660.
 44. **Sun W*** (2012)

- A statistical framework for eQTL mapping using RNA-seq data.
Biometrics. 2012 Mar;68(1):1-11. Epub 2011 Aug 12. PMID: PMC3218220.
45. **Sun W***, Li L (2012)
Multiple loci mapping via model-free variable selection.
Biometrics. 2012 Mar;68(1):12-22. Epub 2011 Aug 12. PMID: PMC3218235.
 46. 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), PMID: PMC3218829.
* co-corresponding author
 47. 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.
 48. 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).
 49. 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. PMID: PMC2947907.
 50. **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. PMID: PMC2870969.
 51. 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. PMID: PMC2815660.
 52. **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
 53. **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. PMID: PMC2935461.
 54. **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). PMID: PMC2700807
55. **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. PMID: PMC2648986.
Co-first authors
 56. 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). PMID: PMC2579309.
Co-first authors
 57. **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). PMID: PMC2432080.
Co-first authors
 58. **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.
 59. 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). PMID: PMC1868765.
 60. 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.
 61. 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. PMID: PMC524832.

Peer Reviewed Publications (Applied-Collaborative)

62. Labadie, J. D., Hua, X., Harrison, T. A., Banbury, B. L., Huyghe, J. R., **Sun, W.**, ... & Newcomb, P. A. (2021).
Genetic Predictors of Severe Skin Toxicity in Patients with Stage III Colon Cancer Treated with Cetuximab: NCCTG N0147 (Alliance).
Cancer Epidemiology and Prevention Biomarkers, 30(2), 404-411.
63. Labadie JD, Hua X, Harrison TA, Banbury BL, Huyghe JR, **Sun W**, Shi Q, Yothers G, Alberts SR, Sinicrope FA, Goldberg RM.
Genetic Predictors of Severe Skin Toxicity in Patients with Stage III Colon Cancer Treated with Cetuximab: NCCTG N0147 (Alliance).
Cancer Epidemiology and Prevention Biomarkers. 2021 Feb 1;30(2):404-11.

64. Zaidi SH, Harrison TA, Phipps AI, Steinfeld R, Trinh QM, Qu C, Banbury BL, Georgeson P, Grasso CS, Giannakis M, Adams JB, ..., **Sun W**, Hudson TJ, Peters U. Landscape of somatic single nucleotide variants and indels in colorectal cancer and impact on survival. *Nature communications*. 2020 Jul 20;11(1):1-2.
65. Labadie JD, Harrison TA, Banbury B, Amtay EL, Bernd S, Brenner H, Buchanan DD, Campbell PT, Cao Y, Chan AT, ... **Sun W**, ..., Newcomb P. Postmenopausal Hormone Therapy and Colorectal Cancer Risk by Molecularly Defined Subtypes and Tumor Location. *JNCI Cancer Spectrum*. 2020 Aug;4(5).
66. Penney KL, Banbury BL, Bien S, Harrison TA, Hua X, Phipps AI, **Sun W**, Song M, Joshi AD, Alberts SR, Allegra CJ. (2019) Genetic Variant Associated with Survival of Patients with Stage II-III Colon Cancer. *Clinical Gastroenterology and Hepatology*. doi.org/10.1016/j.cgh.2019.11.046
67. Grasso CS, Giannakis M, ... , Sun W, ..., Gabriel SB, Lander ES, Garraway LA, Hudson TJ, Fuchs CS, Ribas A, Ogino S, Peters U. (2018) Genetic Mechanisms of Immune Evasion in Colorectal Cancer. *Cancer Discov*. 8(6):730-749.
68. Kim Y, Giusti-Rodriguez P, Crowley JJ, Bryois J, Nonneman RJ, Ryan AK, Quackenbush CR, Iglesias-Ussel MD, Lee PH, **Sun W**, de Villena FP, and PF Sullivan (2018) Comparative genomic evidence for the involvement of schizophrenia risk genes in antipsychotic effects. *Molecular Psychiatry*, 2018 Mar;23(3):708.
69. Mutlu E, Gao L, Collins LB, Walker NJ, Hartwell HJ, Olson JR, **Sun W**, Gold A, Ball LM, Swenberg JA. (2016) Polychlorinated Biphenyls Induce Oxidative DNA Adducts in Female Sprague-Dawley Rats. *Chemical research in toxicology*. 29(8):1335-44.
70. Sharma V, Collins LB, Chen TH, Herr N, Takeda S, **Sun W**, Swenberg JA, Nakamura J. (2016) Oxidative stress at low levels can induce clustered DNA lesions leading to NHEJ mediated mutations. *Oncotarget*. 2016 May 3;7(18):25377-90.
71. **Sun W**[#], Kechris K[#], Jacobson S[#], ..., O'Neal WK, Bowler RP (2016) Common Genetic Polymorphisms Influence Blood Biomarker Measurements in COPD *PLoS Genetics*, 12(8), e1006011.
Co-first authors
72. Jansen R, Penninx BW, Madar V, Xia K, Milaneschi Y, Hottenga JJ, Hammerschlag AR, Beekman A, van der Wee N, Smit JH, Brooks AI, Tischfield J, Posthuma D, Schoevers R, van Grootheest G, Willemsen G, de Geus EJ, Boomsma DI, Wright FA,

- Zou F, **Sun W**, Sullivan PF. (2016)
Gene expression in major depressive disorder.
Mol Psychiatry. 2016 Mar;21(3):339-47.
73. O'Neal WK, Gallins P, Pace RG, Dang H, Wolf WE, Jones LC, Guo X, Zhou YH, Madar V, Huang J, Liang L, Moffatt MF, Cutting GR, Drumm ML, Rommens JM, Strug LJ, **Sun W**, Stonebraker JR, Wright FA, Knowles MR. (2015)
Gene Expression in Transformed Lymphocytes Reveals Variation in Endomembrane and HLA Pathways Modifying Cystic Fibrosis Pulmonary Phenotypes.
Am J Hum Genet. 96 (2), 318-328.
74. Crowley, J. J. #, Zhabotynsky, V. #, **Sun, W.** #, Huang, S., Pakatci, I. K., Kim, Y., ... & Pardo-Manuel, D. V. F. (2015).
Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance.
Nature Genetics, 47(4), 353.
Co-first authors
75. Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., **Sun, W.**, Xia, K., ... & Boomsma, D. I. (2014).
Heritability and genomics of gene expression in peripheral blood.
Nature Genetics, 46(5), 430-437.
76. 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.
77. 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 PMID: PMC3627555
78. 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. PMID: PMC3511858
79. 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. PMID: PMC3513569
80. 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. PMID: PMC3405208

81. **Collaborative Cross Consortium** (2012)
The genome architecture of the Collaborative Cross mouse genetic reference population.
Genetics. 190(2):389-401. PMID: PMC3276630
82. Xia K, Shabalina 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. PMID: PMC3268245.
83. **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. PMID: PMC3284328
84. 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. PMID: PMC3099420
85. 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.
86. 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. PMID: PMC3109544
Co-first authors
87. 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
88. 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. PMID: PMC3296486.
89. 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 (<https://github.com/Sun-lab>)

- asSeq
(Allele-specific) RNA-seq data processing and 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, University of Notre Dame, Oct 27 2014**, 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
21. **National Jewish Health, Denver, Jan 12, 2015**, Title: Genetic basis of gene expression in human and mouse
22. **Department of Biostatistics & Informatics, Colorado School of Public Health, Jan 13, 2015**, Title: Statistical Methods for Cancer Genomics

23. **Bioinformatics Research Center, NC state University, Feb 10, 2015**, Title: Statistical Methods for Cancer Genomics
24. **2015 IMS-China International Conference on Statistics and Probability, Kuming, China, July 1-4**, Title: PenPC: A Two-step Approach to Estimate the Skeletons of High Dimensional Directed Acyclic Graphs
25. **2015 ICSA China Statistics Conference, Shanghai, China, July 6-7**, Title: eQTL mapping and integrative analysis of cancer genomic data
26. **60th World Statistics Congress, Rio de Janeiro, Brazil, July 26-31, 2015**, Title: PenPC: A Two-step Approach to Estimate the Skeletons of High Dimensional Directed Acyclic Graphs
27. **Department of Biostatistics, University of Michigan, Sep 30, 2015**, Title: The interplay of somatic copy number aberration, DNA methylation, and gene expression
28. **Department of Statistics, Oregon State University, May 16, 2016**, Title: Statistical methods to exploit allele-specific and isoform-specific information from RNA seq data
29. **2016 JSM, Chicago, July 30-Aug 4, 2016**, Using Multiple Types of Omic Data to study ITH and beyond
30. **2017 ENAR, Washington DC, March 12-15, 2017** Title: Studying Intra-Tumor Heterogeneity using DNA Sequencing Data from Single Cells and Bulk-Tissue Sample
31. **Biostatistics Branch, DCEG, National Cancer Institute, March 16, 2017**, Title: The interplay of somatic copy number aberration, DNA methylation, and gene expression
32. **Department of Biostatistics, University of Florida, April 7-8, 2017**, Biostatistics workshop: "Statistical Inference for Biomedical Big Data", Title: Mapping Gene Expression QTL of Impure Tumor Samples
33. **14th Graybill Conference on Statistical Genomics and Genetics, June 5-7, 2017**, Title: Analyzing Cancer Omic Data as Compositional Data
34. **2018 ENAR, Atlanta, GA, March 25-28, 2018**, Title: Estimation of Intra-Tumor Heterogeneity and Assessing its Impact on Survival Time
35. **Department of Biostatistics, Ohio State University, November 16, 2018**, Title: Intra-Tumor Heterogeneity and Its Associations with Survival Time and Immunotherapy Treatment
36. **Department of Biostatistics, MD Anderson, March 20th, 2019**, Title: Statistical Analysis of Somatic Mutations in Cancer Genomes
37. **Department of Biostatistics, Duke University, October 11th, 2019**, Title: eQTL Mapping Using RNA-seq Data, assumptions, implementations, and cautions
38. **Department of Biostatistics, University of Chicago, October 23rd, 2019**, Title: CARseq: Cell type Aware analysis of RNA-seq data

39. **iBright Conference 2019, MD Anderson, November 11-13, 2019**, Title: CARseq: Cell type Aware analysis of RNA-seq data
40. **Penn State Statistics Colloquia, Department of Statistics, The Pennsylvania State University, Feb 11, 2021**, Title: Cell type-aware analysis of RNA-seq data,
41. **Biostatistics Research Seminar, Department of Biostatistics, St. Jude Children's Research Hospital, April 27th, 2021**, Title: Cell type-aware analysis of RNA-seq data,
42. **2021 JSM, Virtual meeting, Aug 8-Aug 12, 2021**, Analyzing Genomic Data Using Neural Networks

Teaching (UNC)

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 Post Doctoral Research Fellow

Yang Liu (2016-2018), co-advised with Dr. Qianchuan (Chad) He, Current Position: Assistant Professor, Department of Mathematics and Statistics, Wright State University

Aaron Molstad (2017-2019), co-advised with Dr. Li Hsu, Current Position: Assistant Professor, Department of Statistics, University of Florida

Mengqi Zhang (2019-2020) scRNA-seq data. Current Position: Post doctoral fellow, University of Pennsylvania

Si Liu, (2020-) scRNA-seq data and TCR data

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. Current Position: Analytic Lead at Monsanto Company.

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.

Current Position: Research Assistant Professor at Department of Biostatistics, UNC Chapel Hill

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

Current Position: Assistant Professor at Department of Biostatistics at MD Anderson Cancer Center

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

Current Position: Assistant Professor at Department of Mathematics, Laval University, Canada

Matt Psioda (2013-2016), “Statistical Methods for DNA Methylation Studies and for Bayesian Clinical Trial Design”, co-advised with Dr. Joe Ibrahim.

Current Position: Research Assistant Professor at Department of Biostatistics, UNC Chapel Hill

Jenny Yang (2014-2017), “Model-free estimation of graphic model”.

Doug Wilson (2014-2018), “Statistical Methods for The Deconvolution of Heterogeneous Cell Type Samples using RNA-seq”, co-advised with Dr. Joe Ibrahim

Paul Little (2014-2019), “Intra-tumor Heterogeneity” and cell type-specific eQTL mapping, co-advised with Dr. Danyu Lin

Vasyl Zhabotynsky (2014-2020), “Statistical Model for RNA-seq data”, co-advised with Dr. Fei Zou

Chong Jin (2014-2020), “Intra-tumor Heterogeneity” and Cell type-aware differential expression co-advised with Dr. Danyu Lin

Laura Zhou (2018-), “Deep learning and neoantigen identification”, co-advised with Dr. Fei Zou

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

Yujia Cai (2021), UW Biostatistics, TCR data analysis

Direction of Undergraduate's Research

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

Qiao Xuanyuan (2021), Deep learning methods for medical imaging data.

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	Zhaoyu Yin	Biostatistics	Fei Zou	2014
25	Matthew Weiser	Genetics	Terry Furey	2015
26	Yuying Xie	Genetics	William Valdar & Yufeng Liu	2015
27	Wei Cheng	Computer Science	Wei Wang	2015
28	Chen-Ping Fu	Computer Science	Leonard McMilan	2015
29	Pearl Liu	Biostatistics, UW	Mike Wu	
30	Angela Zhang	Biostatistics, UW	Mike Wu	
31	David Lim	Biostatistics, UNC	Naim Rashid & Joe Ibrahim	

Research Grant

Ongoing Research Support

5 R01 GM105785-02 (Sun)

05/15/2014-12/31/2022

National Institutes of Health

Statistical Methods for RNA-seq Data Analysis

We propose to develop statistical methods and software for RNA-seq data analysis, with emphasis on dissecting the genetic and epigenetic basis of allele-specific and RNA-isoform-specific expression. Role: PI

R21CA224026 (Sun)

06/20/2018 – 05/31/2021

NIH/NCI

Estimation and association analysis of biomarkers for tumor immune microenvironment

We propose to develop statistical methods and software packages to study tumor immune microenvironment, and associate immune cell composition with survival time. Our project will break new ground for epidemiology studies, for example, to enable stratified association analysis for patients with particular tumor immune microenvironment. Role: PI

R01CA222833 (Dai)

07/20/2018 – 06/30/2023

NIH

Statistical Genetics and Genomics for Epidemiologic Research

In this project, we will develop a suite of statistical methodologies for dissecting the multi-faceted role of genetics and genomics in modern epidemiology, and perform innovative analyses in well-characterized extant populations for prostate cancer research. The methodological topics include precision prevention based on individual genetic susceptibility, epigenetic alterations as the interface of the environment and the genome, and causal inference and mediation. Role: Co-Investigator

2 P01 CA142538-06 (Kosorok)

04/1/2015-03/31/2020

National Cancer Institute

Statistical Methods for Cancer Clinical Trials - Project 3: Statistical/Computational Methods for Pharmacogenomics and Individualized Therapy

This research intends to develop novel and high-impact statistical and computational tools for discovering genetic variants associated with interindividual differences in the efficacy and toxicity of cancer medications and for optimizing drug therapy on the basis of each patient's genetic constitution. Role: Subaward PI

R01 GM07335-01A1 (Ibrahim)

7/1/2016 – 6/30/2020

NIH/NIGMS

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: Subaward PI

R01CA189532 (Hsu)

7/1/2015 – 6/30/2024

NIH/NCI

Statistical Methods for Analysis of Tumor Heterogeneity in Genetic Epidemiology

Tumor heterogeneity can lead to different clinical outcomes and understanding the underlying etiology can lead to novel insights into cancer prevention and treatment. Recent developments in

sequencing technologies have made it possible to characterize somatic mutations at great detail in large numbers of tumors, providing a unique opportunity to study tumor heterogeneity. The objective of this application is to develop statistical methods for assessing association of tumor heterogeneity with clinical outcomes and to identify genetic and environmental risk factors that lead to tumor heterogeneity. Role: Co-Investigator

R01 GM126550 (Liu)
NIH/NIGMS

8/15/2017 – 7/31/2021

Collaborative Research: Unlocking complex co-expression network using graphical models

The goal of this project is to develop biostatistical and machine learning methods to estimate such directed graphical models using gene expression data. For the purpose of precise disease treatment, one needs to account for the heterogeneity of patients (inter-patient heterogeneity) as well as the heterogeneity of cells within a patient (intra-patient heterogeneity) for graphical model estimation. Role: Co-PI

P01 CA 18029 (Appelbaum)
NIH/NCI

5/10/2018 – 4/30/2023

Adult Leukemia Research Center, Sub-Project: Core B (Gooley)

The goal of this Program Project Grant is to develop improved, curative therapies for patients with hematological malignancies focusing primarily on adoptive cell therapy and hematopoietic cell transplantation. Individual Projects will explore the clinical use of genetically engineered T cells to treat acute myeloid leukemia and multiple myeloma, and will define the type of graft-versus-host disease prophylaxis that provides the greatest likelihood for survival without either relapse or graft-versus host disease. The Core provides service in biostatistics/bioinformatics, computing, and data collection to ensure that data generated will address appropriately the scientific question at hand. Role: Co-investigator.

U01 ES029516 (Bielas, Li)
NIH/NIEHS

7/1/2018 – 6/30/2023

PREcision Disease preventIon via somatiC muTagenesis enumeratION (PREDICTION)

The overarching goals of this proposal are to characterize the spectrum and burden of spontaneous and induced somatic DNA mutations in humans at unprecedented depth and fidelity, and to evaluate the utility of mutagenesis as a marker of exposure and predictor of risk for developing a disease. Role: Co-investigator, with zero percent effort for years 1-3.

Completed Research Support

U01CA137088 (Peters)
NIH/NCI

9/1/2014 – 8/31/2020

Molecular Pathological Epidemiology of Colorectal Cancer (GECCO)

Role: Co-Investigator

R01 CA176272 (Newcomb/Chan)
National Cancer Institute

4/22/2013 – 9/30/2018

Molecular Correlates of Outcomes in Clinical Trials of Colon Cancer

Role: Co-Investigator

5 R01 CA082659-16 (Lin)
National Cancer Institute

4/1/2000 - 1/31/2017

Statistical Methods in Chronic Disease Research

Role: Co-Investigator

5 R01 GM074175-08 (Zou) 4/1/2006 - 8/31/2015
National Institute of General Medical Sciences
Robust Methods for Complex Trait Mapping with Collaborative Cross
Role: Co-Investigator

5 R01 HG006292-05 (Li) 8/23/2011 - 5/31/2016
National Human Genome Research Institute
Design and Analysis of Sequencing-based Studies for Complex Human Traits
Role: Co-Investigator

5 R01 MH101819-02 (Wright) 8/15/2013- 7/31/2016
North Carolina State University
Systems Approaches to Link Tissue-Specific Expression to Disease
Role: Co-Investigator

5 P30 ES010126-14 (Swenberg) 2/1/2000 - 3/31/2015
National Institute of Environmental Health Sciences
UNC-CH Center for Environmental Health & Susceptibility - Facility Core 2: Biostatistics & Bioinformatics
Role: Core Director

5 R01 CA149569-05 (Liu) 2/1/2010 - 12/31/2015
National Cancer Institute
Flexible Statistical Machine Learning Techniques for Cancer-Related Data
Role: Co-Investigator

5 U01 HL096899-04 (Couper) 7/7/2010-4/30/2015
National Heart, Lung and Blood Institute
ARIC Neurocognitive Study (ARIC-NCS)
Role: Co-Investigator