

NANCY RUONAN ZHANG

SEPTEMBER 26, 2011

CONTACT INFORMATION

Mailing address Department of Statistics
The Wharton School, University of Pennsylvania
400 Jon M. Huntsman Hall
3730 Walnut Street
Philadelphia, PA 19104-6340

E-mail nzh@wharton.upenn.edu

Website <http://www-stat.stanford.edu/~nzhang/>

EMPLOYMENT HISTORY

<i>Associate Professor</i> <i>(with tenure)</i>	7/2011 - current	Dept. of Statistics The Wharton School University of Pennsylvania
<i>Associate Professor</i> <i>(with tenure)</i>	6/2012	Dept. of Statistics Stanford University, CA.
<i>Assistant Professor</i>	9/2006 - 6/2012	Dept. of Statistics Stanford University, CA.
<i>Post-doctoral Fellow</i>	10/2005 - 7/2006	Depts. of Statistics and Plant Biology University of California at Berkeley

EDUCATION

<i>Doctor of Philosophy</i>	9/2005	Dept. of Statistics Stanford University, CA.
<i>Master of Science</i>	6/2001	Dept. of Computer Science Stanford University, CA.
<i>Bachelor of Science</i>	6/2001	Dept. of Mathematics Stanford University, CA

PUBLICATIONS

Articles in refereed Journals:

- (1) Zhang, N.R. and Siegmund, D.O., 2007, A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* 63, 22–32.
- (2) Chan, H.P. and Zhang, N.R., 2007, Scan statistics with weighted observations. *JASA Theory and Methods*, 102, 595–602.*
- (3) The Encode Consortium, 2007, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.
- (4) Lai T.L., Xing, H and Zhang, N.R., 2008, Stochastic segmentation models for array-based comparative genomic hybridization data analysis. *Biostatistics*, 9, 290-307.*

- (5) Zhang, N.R., Wildermuth, M.C. and Speed, T.P., 2008, Transcription factor binding site prediction with multivariate gene expression data. *Annals of Applied Statistics*, 2, 332–365.
- (6) Zhang, N.R., Senbabaoglu, Y. and Li, J., 2010, Joint estimation of DNA copy number from multiple platforms. *Bioinformatics*, 26, 153–160.
- (7) Siegmund, D.O., Yakir, B. and Zhang, N.R., 2010, Tail approximations for maxima of random fields by likelihood ratio transformations. *Sequential Analysis*, 29, 245–262.*
- (8) Zhang, N.R., Siegmund, D.O., Ji, H., and Li, J., 2010, Detecting simultaneous change-points in multiple sequences. *Biometrika*, 97, 631–645.
- (9) Li, F. and Zhang, N., 2010, Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association – Theory and Methods*, 105, 1202-1214.*
- (10) Bickel, P., Boley, N., Brown, B., Huang, H. and Zhang, N.R., 2010, Subsampling methods for genomic inference. *Annals of Applied Statistics*, 4, 1660-1697.*
- (11) Chan, H.P.[‡], Zhang, N.R.[‡], and Chen, L.H.S., 2010, Importance sampling of word patterns in DNA and protein sequences. *Journal of Computational Biology*, 17, 1697-1709.
- (12) Chen, H., Xing, H., and Zhang, N.R.[†], 2011, Estimation of parent specific DNA copy number in tumors using high-density genotyping arrays, *PLoS Computational Biology*, 7, e1001060.
- (13) Siegmund, D.O., Yakir, B. and Zhang, N.R., 2011, Detecting simultaneous variant intervals in aligned sequences. *Annals of Applied Statistics*, 5, 645-668.*
- (14) Efron, B. and Zhang, N.R., 2011, False Discovery Rates and Copy Number Variation, *Biometrika*, 98, 251-271.*
- (15) Natsoulis, G., Bell, J.M., Xu, H., Buenrostro, J.D., Ordonez, H., Grimes, S., Newburger, D., Jensen, M., Zahn, J.M., Zhang, N. and Hanlee P. Ji, 2011, A Flexible Approach for Highly Multiplexed Candidate Gene Targeted Resequencing. *PLoS One*. 2011; 6(6):e21088.
- (16) Siegmund, D., Yakir, B. and Zhang, N.R., 2011, The false discovery rate for scan statistics, *Biometrika*, 98, 979-985.*
- (17) Zhang, N.R. and Siegmund, D., 2011, Model Selection for High Dimensional, Multi-sequence Change-point Problems, *Statistica Sinica*, 22, 1507-1538.
- (18) Muralidharan, O., Natsoulis, G., Bell, J., Newburger, D., Xu, H., Keta, I., Ji, H. and Zhang, N.[†], 2011, A Cross-Sample Statistical Model for SNP Detection in Short-Read Sequencing Data. *Nucleic Acids Research*, 40, e5.

- (19) Flaherty P., Natsoulis G., Muralidharan O., Winters M., Buenrostro J., Bell J., Brown S., Holodniy M., Zhang N., Ji H.P., 2011, Ultra-sensitive detection of rare mutations using next-generation targeted resequencing. *Nucleic Acids Research*, 40, e2.
- (20) Shen, J. and Zhang, N.R.[†], 2011, Change-point model on nonhomogeneous Poisson processes with application in copy number profiling by next-generation DNA sequencing. *Annals of Applied Statistics*, 6, 476-496.
- (21) Muralidharan, O., Natsoulis, G. and Zhang, N.[†], 2011, Detecting mutations in mixed sample sequencing data using empirical Bayes. *Annals of Applied Statistics*, 6, 1047-1067.
- (22) Sun, Y., Zhang, N.R. and Owen, A., 2012, Multiple hypothesis testing, adjusting for latent variables. *Annals of Applied Statistics*, 6, 1664-1688.
- (23) Chen, H. and Zhang, N.R., 2013, Graph-based tests for two-sample comparisons of categorical data. *Statistica Sinica* 23: 1479-1503.
- (24) Natsoulis G., Zhang N., Welch K., Bell J., and Ji H.P., 2013, Identification of insertion deletion mutations from deep targeted resequencing. *Journal of Datamining in Genomics and Proteomics* 4: 132.
- (25) Chen, H., Bell, J. M., Zavala, N. A., Ji, H. P. and Zhang, N. R.[†], 2014. Allele-specific copy number profiling by next-generation DNA sequencing. *Nucleic Acids Research*, 43, e23.
- (26) Nadauld, L. D., Garcia, S., Natsoulis, G., Bell, J. M., Miotke, L., Hopmans, E. S., Xu, H., Pai, R. K., Palm, C., Regan, J. F., Chen, H., Flaherty, P., Ootani, A., Zhang, N. R., Ford, J. M., Kuo, C. J. and Ji, H. P. 2014. Metastatic tumor evolution and organoid modeling implicate TGFBR2 as a cancer driver in diffuse gastric cancer. *Genome Biology* 15:428.
- (27) Chen, H. and Zhang N.[†], 2015, Graph-based change-point detection. *The Annals of Statistics* 43:139.
- (28) Jiang, Y., Oldridge, D. A., Diskin, S. J. and Zhang, N.R.[†], 2015, CODEX: a normalization and copy number variation detection method for whole exome sequencing. *Nucleic Acids Research*, 43, e39.
- (29) Cushing, A., Kamali, A., Winters, M., Hopmans, E.S., Bell, J.M., Grimes, S.M., Li, C.X., Zhang, N.R., Moss, R.B., Holodniy, M., Ji, H.P., 2015. Emergence of Hemagglutinin Mutations During the Course of Influenza Infection, *Science Reports*. 5, 16178.
- (30) Peixoto LL, Wimmer ME, Poplawski SG, Tudor JC, Kenworthy CA, Liu S, Mizuno K, Garcia BA, Zhang NR, Giese K, Abel T. 2015, Memory acquisition and retrieval impact different epigenetic processes that regulate gene expression, *BMC Genomics*, 16, S5.

- (31) Yue, M., Han, X., De Masi, L., Zhu, C., Ma, X., Zhang, J., Wu, R., Schmieder, R., Kaushik, R.S., Fraser, G.P., Zhao, S., McDermott, P.F., Weill, F.X., Mainil, J.G., Arze, C., Fricke, W.F., Edwards, R.A., Brisson, D., Zhang, N.R., Rankin, S.C., Schifferli, D.M., 2015. Allelic variation contributes to bacterial host specificity, *Nature Communications*, 6, 8754.
- (32) Wang, X., Chen, M., Yu, X., Pornputtpong, N., Chen, H., Zhang, N.R., Powers, R.S., Krauthammer, M., 2016, Global copy number profiling of cancer genomes, *Bioinformatics*, 32, 926-8.
- (33) Zhang, N.R., Yakir, B., Xia, L.C., and Siegmund, D.O., 2016, Scanning a Poisson Random Field for Local Signals, *Annals of Applied Statistics*, 10, 726-755.
- (34) Xia, L.C., Sakshuwong, S., Hopmans, E.S., Bell, J.M., Grimes, S.M., Siegmund, D.O., Ji, H.P., and Zhang, N.R.[†], 2016, A genome-wide approach for detecting novel insertion-deletion variants of mid-range size, *Nucleic Acids Research*, *first published online June 20, 2016 doi:10.1093/nar/gkw481*.
- (35) Jiang Y, Qiu Y, Minn AJ, Zhang N.R.[†], 2016, Assessing tumor heterogeneity and tracking longitudinal and spatial clonal evolution by next-generation sequencing. *Proceedings of the National Academy of Sciences*. 113, E5528-E5537.
- (36) Wang, X., Chen, H., Zhang, N.R., 2017, DNA copy number profiling using single-cell sequencing, *Briefings in Bioinformatics*, *accepted*.
- (37) Jiang Y, Zhang N.R.[†], Li M.[†], 2017, Modeling transcriptional bursting and allele-specific expression by single-cell RNA-sequencing, *Genome Biology*, *accepted*.
- (38) Chen, H., Jiang, Y., Maxwell, K., Nathanson, K., Zhang N.R.[†], 2017, Allele-specific copy number estimation by whole exome sequencing, *Annals of Applied Statistics*, *accepted*.
- (39) Jia, C., Li, M.[†], and Zhang, N.R.[†], 2017, Accounting for technical batch effects in single-cell RNA sequencing analysis, *under third round review for Nucleic Acids Research*.
- (40) Jiang Y, Zhang N.R.[†], 2017, Common copy number variation force-calling by whole-exome sequencing, *Submitted*.

Published book chapters:

- (41) Chan, H.P., Tu, I.P. and Zhang, N.R., 2009, Boundary crossing probability computations in the analysis of scan statistics. in *Scan Statistics - Methods and Applications*. Birkhauser, Boston.*
- (42) Zhang, N.R., 2010, DNA copy number profiling in normal and tumor genomes. *Frontiers in Computational and Systems Biology*, ed. Jianfeng Feng, Wenjiang Fu and Fengzhu Sun. pp. 259–281.

* Authors are listed in alphabetical order.

- † Corresponding or co-corresponding author.
‡ Joint first authors.

TEACHING

I have taught the following courses:

- 06-07 Winter, 07-08 Winter:** Applied Statistics (Stanford STAT191)
08-09 Winter, 09-10 Winter: Introduction to ANOVA (Stanford STAT203)
06-07 Spring, 07-08 Spring: Nonparametric Statistics (Stanford STAT205)
08-09 Winter, 09-10 Winter, 10 Fall: Stochastic Processes with Applications in Biology (Stanford STAT215)
08-09 Spring: Computational Algorithms in Statistical Genetics (Stanford STAT345/GEN245)
09-10 Spring: Statistical Methods in Genetics (Stanford STAT166/STAT366)
11-12 Spring: Introductory Statistics (Wharton STAT102)
12-13 Fall: Introductory Statistics (Wharton STAT431)
12-13 Spring: Intermediate Statistics (Wharton STAT471)
14-15 Spring: Introductory Statistics (Wharton STAT102)
16-17 Fall: Statistical Computing with R (Wharton STAT405/705)

HONORS AND GRANT AWARDS

- 2014-2017:** NIH 2-R01-HG-006137, co-PI, Project title: “Statistical Models and Analysis of Complex Genomic Variation in Clonal Mixtures”
2016-2020: NSF-NIGMS Grant, co-PI, Project title: “Statistical Methods for High-Resolution Multiscale Analysis 3D DNA”
2011: Sloan Fellowship, Sloan Foundation
2011-2014: NIH R01 HG006137, PI, Project title: “Statistical Models for Genome Sequencing and Association”.
2010-2013: NSF DMS Grant ID 1043204, co-PI with David Siegmund, Project title: “Statistical Methods for Threat Detection”
2009-2012: NSF DMS Grant ID 0906394, PI, Project title: “Change-point Problems in Genomic Profiling”
2006-current: Terman Fellowship, Stanford University
2007: New World Silver Medal for the Best PhD Thesis in the Mathematical Sciences
2002: Honorary Stanford Graduate Fellow
2002: National Defense Science and Engineering Graduate Fellowship

PAST AND CURRENT DOCTORAL STUDENTS AND POST-DOCTORAL FELLOWS

- (1) Current: Jingshu Wang (Post-doc), Department of Statistics, The Wharton School
- (2) Current: Li Xia (Post-doc), Department of Medicine, Stanford University (Joint with Hanlee Ji).

- (3) Current: Xuran Wang, Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania (Joint with Dylan Small).
- (4) Current: Yuchao Jiang, Graduate Group in Genomics and Computational Biology, University of Pennsylvania, graduating Spring 2016.
- (5) 2014: Hao Chen, Department of Statistics, Stanford University, currently in Department of Statistics faculty of Univ. of California, Davis.
- (6) 2012: Jeremy Shen, Department of Statistics, Stanford University, currently at Two Sigma Investments.
- (7) 2012: Yunting Sun, Department of Statistics, Stanford University (Joint with Art Owen), currently at Google Inc..

SERVICE AND SYNERGISTIC ACTIVITIES

- 2014-current:** Doctoral program co-advisor for Graduate Group in Genomics and Computational Biology.
- 2012-current:** Doctoral program co-director for Wharton Statistics Department.
- 2010 fall:** Masters student advisor for Stanford statistics department.
- 2007-current:** Faculty advisor for the Computational Mathematics Undergraduate major at Stanford University.
- 2006-current:** Part of the development team for the statistical analysis portal for ENCODE data.
- 2009 spring:** Organized the workshop titled “Mathematical Genomics” during April 13, 2009 to April 15, 2009 at the Mathematical Sciences Research Institute in Berkeley.
- 2007 summer:** Coordinated the Stanford VPUE Undergraduate Research Program, advised undergraduates in their research projects.

INVITED TALKS

- 2006/01:** IPAM Workshop in Mathematical Biology, UCLA
- 2006/01:** Statistics Department, Stanford University
- 2006/02:** Statistics Department, UC Berkeley
- 2006/07:** Statistica Sinica, Taiwan
- 2006/07:** National Health Research Institute, Taiwan
- 2006/11:** Department of Epidemiology and Biostatistics, UC San Francisco
- 2007/04:** Biostatistics Workshop, Stanford University
- 2007/05:** Bayesian Statistics Conference, Harvard University
- 2007/06:** Graybill Conference, Fort Collins, Colorado
- 2007/06:** Taipei International Statistical Symposium, Taiwan

- 2007/08:** International Conference on the Frontiers of Statistics: High Dimensional Data Analysis, Kunming, China
- 2007/09:** Department of Statistics, Duke University
- 2007/10:** Department of Biomathematics, UCLA
- 2008/01:** Computational Biology Seminar, Carnegie Mellon University
- 2008/04:** Biostatistics Workshop, Stanford University
- 2008/06:** WNAR, Davis, CA
- 2008/07:** International Workshop on Applied Probability, Campiegne, France
- 2008/08:** Joint Statistical Meetings, Denver
- 2008/09:** Workshop Change-Point Detection Methods and Applications, Paris
- 2008/10:** Department of Statistics, UC Berkeley
- 2008/11:** Department of Epidemiology and Biostatistics, UC San Francisco
- 2009/04:** Department of Statistics, University of Chicago
- 2009/04:** Mathematical Biology Workshop, Mathematical Sciences Research Institute, Berkeley
- 2009/06:** Conference in Honor of 70-th Birthday of Minping Qian
- 2009/06:** ICSA Applied Statistics Symposium, San Francisco, CA
- 2009/06:** IMS Asian Pacific Rim Meeting, Seoul, Korea
- 2009/06:** Statistical Genomics Workshop, Singapore
- 2009/07:** Department of Mathematics, Nankai University, Tianjin, China
- 2009/07:** IMS-China International Conference on Statistics and Probability
- 2009/10:** Department of Statistics, UC Davis
- 2009/10:** Seminar in Computational Biology, University of Southern California
- 2009/10:** Department of Statistics, Stanford University
- 2010/01:** Neyman Seminar, UC Berkeley
- 2010/02:** Seminar in Computational Biology, Stanford University
- 2010/06:** ICSA Applied Statistics Symposium, Indianapolis (Jeremy Shen gave the talk.)
- 2010/08:** Joint Statistical Meetings, Vancouver, Canada
- 2010/12:** Department of Statistics, University of Pennsylvania
- 2011/03:** ENAR, Miami, FL (Jeremy Shen gave the talk.)
- 2012/04:** Department of Electrical Engineering, University of Delaware
- 2012/05:** Conference on Next Generation Sequencing Data, Iowa State University, IA
- 2012/06:** ICSA, Boston, MA
- 2012/08:** JSM, San Diego, CA
- 2013/04:** Department of Statistics, Harvard University