

## CURRICULUM VITAE

### TIAN ZHENG

Department of Statistics  
Columbia University  
Room 1005, MC 4690  
1255 Amsterdam Ave  
New York, New York 10027  
Phone: (212) 851-2134  
Fax: (212) 851-2164  
Email: [tzheng@stat.columbia.edu](mailto:tzheng@stat.columbia.edu)  
Homepage: <http://www.stat.columbia.edu/~tzheng>

---

#### EDUCATION

- 09/1998–05/2002 Columbia University, New York, NY  
Ph.D. in Statistics (2002). M.A. in Statistics (2000). Thesis: Multiple-Marker Screening Approach Towards the Study of Complex Traits in Human Genetics. Advisor: Professor Shaw-Hwa Lo.
- 09/1994–07/1998 Tsinghua University, Beijing, China  
B.S. in Applied Mathematics (1998). Secondary Degree in Computer Science (1998). Thesis: Mathematical Modeling and Parametrization of a Class of Consonants.
- 

#### EXPERIENCE

- 07/2007–present Associate professor, Department of Statistics, Columbia University, New York, NY.
- 07/2002–06/2007 Assistant professor, Department of Statistics, Columbia University, New York, NY.
- 05/2001–08/2001 Statistics consultant, Non-clinical biostatistics, R.W.Johnson pharmaceutical research institute, Raritan, NJ.
- 2001–2002 Research assistant, Department of Statistics, Columbia University
- 2000–2001 Graduate student instructor, Department of Statistics, Columbia University
- 1999–2000 Teaching assistant, Department of Statistics, Columbia University
- 

#### RESEARCH INTERESTS

**Statistical genetics:** Association mapping of complex traits.

**Bioinformatics and Computational genomics:** Feature selection and classification for gene expression data, data mining of microarray data, hidden Markovian models of DNA mutations and phylogenies inference, evaluation of protein regulatory network.

**Statistics:** Bootstrap methods, nonparametric methods, modeling of (social) network structure, design and analysis of sample surveys, statistical education.

---

HONORS

- 2009 Winner of JSM 2009 Poster Award, Section on Survey Research Methods (SRMS) of the American Statistical Association
- 2008 The Mitchell Prize, Section on Bayesian Statistical Science (SBSS) of the American Statistical Association and the International Society for Bayesian Analysis (ISBA)
- 2008 Outstanding Statistical Application Award, American Statistical Association
- 2007 Achievement award, WorldComp'07 (The 2007 World Congress in Computer Science, Computer Engineering and Applied Computing).
- 2007 Worldcomp'07 honorary first runner-up for best paper award in Bioinformatics and Computational Biology.
- 1998–2002 Faculty fellowship, Columbia University.
- 1998 Graduate of honors with distinction (awarded to top 2% of the graduating class), Tsinghua University.
- 1996–1997 Oriental scholarship for top students, Tsinghua University.
- 1995–1996 “12.9” scholarship for top students, Tsinghua University.
- 1994–1995 Excellent student scholarship, Tsinghua University.
- 

## RESEARCH GRANTS

- 2009–2011 *Computational Discovery of Synergistic Mechanisms Responsible for Psychiatric Disorders*, PI: D Anastassiou and M Karayiorgou.  
National Institutes of Health, (Role: co-investigator, one-month effort.)
- 2009–2011 ARRA Administrative supplement to “*New Approach Towards Mapping of Complex Human Disorders*”, PI: S-H Lo.  
National Institutes of Health, (Role: co-PI, two-month effort.)
- 2007–2010 *A General Framework for High Throughput Biological Learning: Theory Development and Applications*, PI: S-H Lo.  
National Science Foundation (Role: co-PI, one-month effort.)
- 2006–2010 *Computer System for Functional Analysis of Genomic Data*, PI: A Rzhetsky.  
National Institutes of Health R01 grant, (Role: co-investigator, 10% effort.)
- 2005–2010 *New Approach Towards Mapping of Complex Human Disorders*, PI: S-H Lo.  
National Institutes of Health R01 grant, (Role: co-PI, 16.7% effort.)
- 2005–2008 *Design and Analysis of “How many X’s do you know” surveys for the study of polarization in social networks*, PI: A Gelman.  
National Science Foundation (Role: co-PI, one-month effort.)

## PEER-REVIEWED PUBLICATIONS

1. Lo, S.H. and Zheng, T. (2002) Backward haplotype transmission association (BHTA) algorithm—a fast multiple-marker screening method, *Hum Hered*, **53**, 197-215.
2. Lo, S.H. and Zheng, T. (2004) A demonstration and findings of a statistical approach through reanalysis of inflammatory bowel disease data, *Proc Natl Acad Sci U S A*, **101**, 10386-10391.
3. Zheng, T., Salganik, M.J. and Gelman, A. (2006) How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks, *Journal of the American Statistical Association*, **101**, 409-423.
4. Zheng, T., Wang, H. and Lo, S.H. (2006) Backward genotype-trait association (BGTA)-based dissection of complex traits in case-control designs, *Hum Hered*, **62**, 196-212.
5. Rzhetsky, A., Zheng, T.<sup>1</sup> and Weinreb, C. (2006) Self-correcting maps of molecular pathways, *PLoS One*, **1**, e61.
6. Zheng, T. and Lo, S.H. (2007) A modified Kendall rank-order association test for evaluating the repeatability of two studies with a large number of objects. In Nair, V. (ed), *Advances in Statistical Modeling and Inference—Essays in Honor of Kjell A. Doksum*. World Scientific, Hackensack, NJ.
7. Kerman, J., Gelman, A., Zheng, T. and Ding, Y. (2007) Visualization in Bayesian data analysis. In Chen, C.-H., Hardle, W. and Unwin, A. (eds), *Handbook of Computational Statistics (Volume III) Data Visualization*. Springer-Verlag, Heidelberg.
8. Zheng, T., Ichiba, T. and Morton, B.R. (2007) Assessing substitution variation across sites in grass chloroplast DNA, *Journal of Molecular Evolution*, **64**, 605-613.
9. Rzhetsky, A., Wajngurt, D., Park, N. and Zheng, T. (2007) Probing genetic overlap among complex human phenotypes, *Proc Natl Acad Sci U S A* **104**, 11694-11699.
10. Ding, Y., Cong, L., Ionita-Laza, I., Lo, S.H. and Zheng, T.<sup>2</sup> (2007) Constructing gene association network for rheumatoid arthritis using the backward genotype-trait association (BGTA) Algorithm. In “Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci.” *BMC Proceedings* **1** S1:S13.
11. Li, Z., Zheng, T., Califano, A. and Floratos, A. (2007) Pattern-based mining strategy for multi-Locus association analysis on GAW15 problem 2 chromosome 18 dataset. In “Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci.” *BMC Proceedings* **1** S1:S16.
12. Wang, S., Zheng, T. and Wang, Y. (2007) Transcription activity hotspot, is it real or an artifact? In “Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci.” *BMC Proceedings* **1** S1:S94.
13. Zheng, T., Wang, S., Cong, L., Ding, Y., Ionita-Laza, I. and Lo, S.H. (2007) Joint study of genetic regulators for expression traits related to breast cancer. In “Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci.” *BMC Proceedings* **1** S1:S10.
14. Yan, X. and Zheng, T.<sup>3</sup> (2007) Discriminant Analysis using multigene profiles in classification of breast cancer. *Proceedings of the 2007 International Conference on Bioinformatics and Computational Biology*. (Acceptance rate: 30%. Worldcomp’07 Honorary First Runner-Up for Best Paper Award in Bioinformatics and Computational Biology.)

---

<sup>1</sup>Joint first author.<sup>2</sup>Correspondence author.<sup>3</sup>Correspondence author.

15. Woo, J.H., Zheng, T.<sup>4</sup> and Kim, J.H. (2007) Identifying genomic regulators of set-wise co-expression. *Proceedings of IEEE 7th International Conference on Bioinformatics and BioEngineering*. (Acceptance for regular research papers is 65 out of  $\sim 500$ .)  
(<http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=4375521&isYear=2007>)  
Under consideration for journal publication by BIBE07 program committee.
16. Chen, G. K., Zheng, T., Witte, J. S. and Goode, E. L. (2007) Genome-wide association analyses of expression phenotypes. *Genetic Epidemiology* **31**(S1):S7-S11.
17. Yan, X. and Zheng, T.<sup>5</sup> (2008) Selecting informative genes for discriminant analysis using multigene expression profiles. *BMC Genomics* **9**(S2):S14.
18. Watkinson, J., Wang, X., Zheng, T. and Anastassiou, D. (2008) Identification of gene-gene interactions associated with disease from gene expression data. *BMC Systems Biology* **2**:10.
19. Iossifov, I., Zheng, T., Baron, M., Gilliam, T.C. and Rzhetsky, A. (2008) Genetic-linkage mapping of complex hereditary disorders to a whole-genome molecular-interaction network. *Genome Research* **18**:1150-1162.
20. Lo, S.-H., Chernoff, H., Cong, L., Ding, Y. and Zheng, T. (2008) Discovering interactions among BRCA1 and other candidate genes involved in Sporadic Breast Cancer. *Proc Natl Acad Sci U S A* **105**, 12387-12392.
21. Watkinson, J., Liang, K.C., Wang, X., Zheng, T., and Anastassiou, D. (2008) Inference of regulatory gene interactions from expression data using three-way mutual information. *The Annals of New York Academy of Sciences* **1158**:302-313.
22. Salicru, M., Vives, S. and Zheng, T.<sup>6</sup> (2008) Inferential clustering approach for two-color spotted microarray experiments. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 13 Oct. 2008. IEEE computer Society Digital Library. IEEE Computer Society,  
<<http://doi.ieeeecomputersociety.org/10.1109/TCBB.2008.106>>
23. Zheng, T. and Gastwirth, J. L. (2008) On Bootstrap Tests of Symmetry About an Unknown Median. *Journal of Data Science*. In press.
24. Huang, C.-H., Cong, L., Xie, J., Qiao, B., Lo, S.-H., Zheng, T.<sup>7</sup> (2008) Rheumatoid Arthritis-Associated Gene-Gene Interaction Network for Rheumatoid Arthritis Candidate Genes. *BMC proceedings for GAW 16*. Accepted.
25. Qiao, B., Huang, C.-H., Cong, L., Xie, J., Lo, S.-H., Zheng, T.<sup>8</sup> (2008) Genome-Wide Gene-Based Analysis of Rheumatoid Arthritis-Associated Interaction with PTPN22 and HLA-DRB1. *BMC proceedings for GAW 16*. Accepted.
26. McCormick, T. H., Salganik, M. J., and Zheng, T. (2008) Efficiently estimating personal network size. *Journal of the American Statistical Association*. In press.
27. Chernoff, H., Lo, S. H. and Zheng, T. (2009) Discovering influential variables: a method of partitions. *Annals of Applied Statistics*. Accepted.
28. Wang, S., Zheng, T., Chanock, S., Jedrychowski, W. and Perera, F. P. (2009) Methods for detecting interactions between genetic polymorphisms and prenatal environment exposure with a mother-child design. *Genetic Epidemiology*. Accepted.
29. Kim, S., Singh, P., Park, J., Park, S., Friedman, A., Zheng, T., Sun, Q., Lee, Y.-H. and Lee, K. (2008) Fungal photoreceptor MGWC-1-mediated disease suppression in rice. Submitted.

---

<sup>4</sup>Joint first author.

<sup>5</sup>Correspondence author.

<sup>6</sup>Joint correspondence author.

<sup>7</sup>Joint correspondence author.

<sup>8</sup>Joint correspondence author.

30. DiPrete, T. A., McCormick, T. H., Gelman, A., Teitler, J., Zheng, T. (2008) Segregation in social networks based on acquaintanceship and trust. Submitted.

#### OTHER PUBLICATIONS

31. Zheng, T. (2002) Multiple-marker screening approach towards the mapping of complex traits in human genetics. Columbia University Doctoral Dissertation.
32. Ding, Y. and Zheng, T. (2006) Tree-Based Integration of One-versus-Some (OVS) Classifiers for Multiclass Classification, *Proceedings of the joint statistical meeting 2006*. (To be extended and submitted for journal publication.)
33. Rosner, E. and Zheng T (2006) Helpfulness of the genders—a student project. STATS. Forthcoming. (This is a magazine style article on Teaching Statistics)
34. McCormics, T. H. and Zheng, T. (2007) Adjusting for recall bias in “how many X’s do you know?” surveys, *Proceedings of the joint statistical meeting 2007*.
35. Zheng, T. and Lo, S.-H. (2008) Discussion on “Quantifying the Fraction of Missing Information for Hypothesis Testing in Statistical and Genetic Studies by Nicolae, Meng and Kong.” *Statistical Science*. **23**(3):321-324.

#### UNPUBLISHED TECHNICAL REPORTS AND WORKING PAPERS

37. Zheng, T. and Lo, S.-H. (2008) Constructing gene association networks for complex human disorders using the backward genotype-trait association (BGTA) algorithm. Technical report. Department of Statistics, Columbia University.
38. Ding, Y. and Zheng, T. (2007) Feature Selection and Classification Based on K-Nearest Neighbor Patterns. Working paper. Department of Statistics, Columbia University.
39. Ionita-Laza, I., Ding, Y., Cong, L., Lo, S.H. and Zheng, T.<sup>9</sup> (2007) Combined Linkage and Association Analysis of the NARAC Dataset. Technical report. Department of Statistics, Columbia University.
40. Zheng, T., Lo, S.H., Reilly, C. and Blumenthal, M.N. (2005) Genetic Analysis of Asthma Data Using the Backward Haplotype Transmission Association (BHTA) method. Technical report. Department of Statistics, Columbia University.
41. Zheng, T., Wang, H. and Lo, S.H. (2004) Rank-based generalized genotype-trait association algorithm for the mapping of quantitative trait loci. Working paper. Department of Statistics, Columbia University.

#### INVITED SEMINAR TALKS

- |            |  |
|------------|--|
| April 2009 | Department of Statistics, Rutgers University           |
| April 2008 | Department of Statistics, George Washington University |
| Feb 2008   | Division of Biostatistics, New York University         |
| Dec 2007   | Department of Statistics, University of Minnesota      |

<sup>9</sup>Correspondence author.

Nov 2007	Department of Statistics, Purdue University
Oct 2007	Department of Biostatistics, University of Pennsylvania
June 2007	Bioinformatics Division, Department of Automation, Institute of Information Processing, Tsinghua University (Beijing, China)
November 2006	Department of Statistics, Yale University.
November 2006	Department of Statistics, Harvard University.
November 2005	University seminar on Genetic Epidemiology, Columbia University.
April 2005	Department of Biostatistics, Yale University.
November 2004	Department of Statistics, Columbia University.
November 2003	Department of Biostatistics, University of Alabama at Birmingham.
October 2002	International Center for Health Outcomes and Innovation Research, Columbia University.
February 2002	Department of Statistics, University of Wisconsin at Madison.
February 2002	Department of Statistics, University of California at Riverside.
February 2002	Department of Statistics, University of California at Davis.
February 2002	Department of Statistics, University of North Carolina at Chapel Hill.
February 2002	Department of Statistics, Pennsylvania State University.
February 2002	Department of Statistics, Columbia University.
February 2002	Department of Statistics, Ohio State University.

---

## CONFERENCE PRESENTATIONS

August 2009	<b>(Winner of poster award)</b> “Latent Structural Models for Aggregated Relational Data”, Joint Statistical Meeting 2009, Washington, DC.
July 2009	<b>(Invited)</b> “Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm,” IMS-APRM 1st meeting, Seoul, Korea.
June 2009	<b>(Invited)</b> “Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm, Workshop on Detecting Influential Variables in High-Dimensional Data, Taida Institute of Mathematical Science, National Taiwan University, Taipei, Taiwan.
Nov 2008	<b>(Peer reviewed)</b> “Immunohistochemical evaluation of uteri from female rhesus monkeys ( <i>Macaca Mulatta</i> ) after cyclic intramuscular estrogen treatment,” American Society of Reproductive Medicine 64th Annual Meeting. (As co-author)
June 2008	<b>(Invited)</b> “Feature selection and classification based on k-nearest-neighbor patterns,” Beijing International Conference on Machine Learning and Data Mining.
June 2008	<b>(Invited)</b> “Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm,” ICSA 2008 applied statistics symposium.
Oct 2007	<b>(Invited)</b> “Studying co-regulation and inter-regulation of genes via eQTL mapping,” IEEE 2007 International Conference on Bioinformatics and BioEngineering.

Oct 2007	<b>(Invited)</b> “Evaluating the repeatability of two studies of a large number of objects: modified Kendall rank-order association test,” Current and Future Trends in Nonparametrics.
August 2007	<b>(Invited)</b> “Studying co-regulation and inter-regulation of genes via eQTL mapping,” Joint Statistical Meeting 2007
June 2007	<b>(Peer-reviewed)</b> “Discriminant analysis using multigene expression profiles in classification of breast Cancer,” BioComp’07.
June 2007	<b>(Invited)</b> “Studying co-regulation and inter-regulation of genes via eQTL mapping,” NSF Sponsored International Conference on Bioinformatics, Hangzhou, China.
August 2006	<b>(Invited)</b> Discussion on “Statistical models for networks,” Joint Statistical Meeting 2006.
August 2006	<b>(Invited)</b> “Design and analysis of ‘how many X’ surveys,” Joint Statistical Meeting 2006.
June 2006	<b>(Invited)</b> “Evaluating the repeatability of two studies of a large number of objects: modified Kendall rank-order association test,” ICSA 2006 applied statistics symposium.
August 2005	“A nonparametric multipoint screening method for QTL mapping,” Joint Statistical Meeting 2005.
June 2005	<b>(Invited)</b> “A nonparametric multipoint screening method for QTL mapping,” International Conference on Statistics in Honour of Professor Kai-Tai Fang’s 65th Birthday.
March 2004	<b>(Invited)</b> “Information-driven marker selection for large scale genomic studies on complex traits,” International Biometric Society, Eastern North American Region (ENAR), Spring Meeting 2004.
August 2003	“Detecting genetic association in case-control studies using backward genotype transmission association (BGTA),” Joint Statistical Meeting 2003.
August 2002	“Haplotype transmission association—an ‘importance’ measure for selecting genetic markers,” Joint Statistics Meeting 2002.

## COURSES TAUGHT

STAT W1111	Introduction to statistics—an introductory level course (aver. enroll.: 120 students).
STAT W4315	Linear Regression Models—core requirement for M.A. in statistics program (aver. enroll.: 65 students).
STAT W8325	Topics in Advanced Statistics—modern statistical learning and computing, with applications to biology, a topic course for PHD students.
STAT W6103	Statistical Modeling/Data Analysis III (core curriculum for PHD program).

## DOCTORAL STUDENTS SUPERVISED

2004	Xin Yan, Department of Statistics, Columbia University
2004	Hui Wang, Department of Statistics, Columbia University
2008	Yuejing Ding, Department of Statistics, Columbia University
In progress	Tyler McCormick, Department of Statistics, Columbia University

---

CONSULTING PROJECTS

May 2008	Consultant to Dr. Jian Tang at Mount Sinai School of Medicine.
2007–present	Consultant to Dr. Andrew Joe at the Department of Medicine, The Medical Center of Columbia University on Microarray data analysis.
2005–present	Consultant to Dr. Doris Chang at the Department of Psychology, New School University on “Chinese immigrant women: domestic abuse and help seeking.”
Summer 2004	Consultant to Dr. Kwangwon Lee at the Department of Plant Pathology, Cornell University on “Characterization of the roles of light in plant-microbe interactions.”
Summer 2003	Consultant to Dr. Steven Schinke of the School of Social Work, Columbia University on clinical trials evaluating the effects of intervention approaches on alcohol use among adolescents.
Summer 2003	Consultant to Intersystems Inc on the Prevention Program Outcomes Monitoring System survey evaluating science-based prevention programs identified and disseminated by the Center for Substance Abuse Prevention (CSAP) of the Substance Abuse and Mental Health Services Administration (SAMHSA).
2002–2003	Consultant to International Center for Health Outcomes and Innovation Research (INCHOIR), Columbia University on “Randomized Evaluation of Mechanical Assistance for Treatment of Congestive Heart Failure (REMATCH).”

---

## PROFESSIONAL SERVICE

## — Editorial service

2007–present	Associate editor, Journal of the American Statistical Association, Applications and Case Studies.
--------------	---

## — Referee

| BioComp'07 | Behavior Genetics | Bioinformatics | BMC Genomics | Communications in Statistics | Human Heredity | IEEE BIBE 07 | IEEE transactions on signal processing | Journal of Official Statistics | Journal of Statistical Planning and Inference | Journal of the American Statistical Association | PLoS genetics | Statistical Applications in Genetics and Molecular Biology | Statistical Methodology | American Medical Informatics Association Symposium 2005 | Prentice Hall |

## — Conference sessions organizers and chairs

2007	Program committee, IEEE 2007 International Conference on Bioinformatics and BioEngineering.
2007	Organized invited session JSM 2007: <i>Statistical methods for gene regulatory networks</i> .
2006–present	Organized invited sessions for ICSA applied statistics symposium 2006: <i>Recent contributions to nonparametric methodologies in biomedical research</i> ; 2008: <i>Designs and analyses of genome-wide association studies</i> .
2005	Organized topic contributed session JSM 2005: <i>Recent statistical contributions to Genetics and Epidemiology</i> .
2003–present	Chaired sessions for Joint Statistical Meetings: 2003, 2005 and 2007.

— **Departmental service**

- 2007-2008      Director of graduate study (Ph.D. program), Department of Statistics, Columbia University.
- 2006-07, 2002-03   Seminar chair, Department of Statistics, Columbia University.
- 2004–present      Doctoral students mentoring: Yuejing Ding (2004), Jiayi Chen (2005), Xiaoru Wu (2006), Li Song (2006), Tyler McCormick (2007), Department of Statistics, Columbia University.
- 2002–present      Computing committee, committee on curriculum, committee on junior search (2002-2003), committee on PhD admissions, Department of Statistics, Columbia University.

— **Doctoral Defense Committees**

- 2008              (Oral Exam) Xiaoru Wu, Department of Statistics, Columbia University
- 2008              Yuejing Ding, Department of Statistics, Columbia University
- 2008              Ivan Iossifov, Department of Medical Informatics, Columbia University
- 2006              (Oral Exam) Yihong Zhao, Department of Biostatistics, Columbia University  
Iuliana Ionita, Department of Computer Science, New York University
- (Oral Exam) Shouhao Zhou, Department of Statistics, Columbia University  
Jouni Kerman, Department of Statistics, Columbia University
- 2005              (Oral Exam) Jouni Kerman, Department of Statistics, Columbia University  
Jing Han, Division of Biostatistics, New York University
- 2004              (Thesis outline defense) Jing Han, Division of Biostatistics, New York University  
Hui Wang, Department of Statistics, Columbia University  
Xin Yan, Department of Statistics, Columbia University
- 2003              Cristian Pasarica, Department of Statistics, Columbia University  
Zaiying Huang, Department of Statistics, Columbia University
- 2002              Yuhang Xing, Department of Finance, Columbia University

**MEMBERSHIPS**

- American Statistical Association
- Royal Statistical Society
- Institute of Mathematical Statistics
- International Chinese Statistical Association