

**Yu Zhang**  
421A Thomas Building  
Department of Statistics  
The Pennsylvania State University  
University Park, PA 16802

[yuzhang@stat.psu.edu](mailto:yuzhang@stat.psu.edu)  
(857) 928-9287 (cell)  
(814) 867-0780 (office)  
(814)-863-7114 (fax)

---

## CURRICULUM VITAE

### RESEARCH INTEREST

Statistics and Computational Biology

### PROFESSIONAL CAREER

2006- Present                      The Pennsylvania State University, Department of Statistics  
Assistant Professor

### POST-DOCTORAL TRAINING

2004-2006                      Harvard University, Department of Statistics

### EDUCATION

2000-2004                      Department of Mathematics, University of Southern  
California, Los Angeles, CA, USA  
Ph.D. in Applied Mathematics

1996-2000                      Department of Physics, Peking University, Beijing, China  
B.S. in Physics

### AWARD AND HONORS

2000-2004                      Graduate Merit Award, University of Southern California  
2001 Summer                  Steed Fellowship, University of Southern California

### PROFESSIONAL MEMBERSHIPS

American Society of Statistics  
American Society of Human Genetics  
International Chinese Statistical Association  
International Society of Computational Biology

### PUBLICATIONS

#### Submitted:

1. **Zhang Y** (2011) A novel Bayesian graphical model for genome-wide multi-SNP association mapping. *Genet Epi*. First revision submitted

#### Accepted (all peer-reviewed):

1. Wu W, Cheng Y, Keller CA, Ernst J, Kumar SA, Mishra T, Morrissey C, Dorman CM, Chen KB, Drautz D, Giardine B, Shibata Y, Song L, Pimkin M, Crawford GE, Furey TS, Kellis M, Miller W, Taylor J, Schuster SC, **Zhang Y**, Chiaromonte F, Blobel GA, Weiss MJ, Hardison RC (2011) Dynamics of the Epigenetic

Landscape During Erythroid Differentiation after GATA1 Restoration. *Genome Res.*

Appeared (all peer-reviewed):

1. **Zhang Y**, Zhang J, Liu JS (2011) Block-Based Bayesian Epistasis Association Mapping with Application to WTCCC Type 1 Diabetes Data. *Ann Appl Stat.* to appear in the Sept issue.
2. **Zhang Y** and Liu JS (2011) Fast and Accurate Approximation to Significance Tests in Genome-Wide Association Studies. *J Am Stat Assoc*, ahead of print. doi:10.1198/jasa.2011.ap10657
3. Miller W, Hayes VM, Ratan A, Petersen DC, Wittekindt NE, Miller J, Walenz B, Knight J, Qi J, Zhao F, Wang Q, Bedoya-Reina OC, Katiyar N, Tomsho LP, Kasson LM, Hardie RA, Woodbridge P, Tindall EA, Bertelsen MF, Dixon D, Pyecroft S, Helgen KM, Lesk AM, Pringle TH, Patterson N, **Zhang Y**, Kreiss A, Woods GM, Jones ME, Schuster SC. (2011) Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil), *Proc Natl Acad Sci.* 108:12348-53
4. Song G, Hsu CH, Riemer C, **Zhang Y**, Kim HL, Hoffmann F, Zhang L, Hardison R, NISC Comparative Sequencing Program, Green E, Miller W (2011). Conversion Events in Gene Cluster. *BMC Evol Biol.* 11:226
5. **Zhang Y**, Jiang B, Zhu J, and Liu JS (2011) Bayesian Models for Detecting Epistatic Interactions from Genetic Data. *Ann Hum Genet.* 75:183-193.
6. **Zhang Y** (2011) Bayesian Epistasis Association Mapping via SNP Imputation. *Biostatistics*, 12:211-222.
7. **Zhang Y**, Book chapter: False Positive Control for Genome-Wide ChIP-Chip Tiling Arrays. (2010) *Frontiers in Computational and System Biology.* Feng J., Fu W, Sun F (eds.), Springer.
8. Chen KB, **Zhang Y** (2010) A Varying Threshold Method for ChIP Peak-Calling Using Multiple Sources of Information. *Bioinformatics*, 26:i504-i510.
9. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, Alkan C, Kidd JM, Sun Y, Drautz DI, Bouffard P, Muzny DM, Reid JG, Wang Q, Nazareth LV, Burhans R, Riemer C, Wittekindt NE, Moorjani P, Tindall EA, Taliwal V, Hubisz MJ, Danko CG, Teo WS, Buboltz A, Zhang Z, Oosthuysen A, Steenkamp A, Oostuisen H, Venter P, Gajewski J, **Zhang Y**, Pugh BF, Makova K, Nekrutenko A, Mardis ER, Patterson N, Pringle T, Chiaromonte F, Mullikin J, Eichler EE, Hardison RC, Gibbs RA, Siepel A, Harkins TT, Hayes VM (2010) Complete Khoisan and Bantu Genomes from Southern Africa. *Nature*, 463: 943-947
10. Hsu CH, **Zhang Y**, Hardison RC, NISC Comparative Sequencing Program, Green ED, and Miller W (2010) An Effective Method for Detecting Gene Conversion Events in Whole Genomes, *J Comp Biol.* 17:1281-1297.
11. Ratan A, **Zhang Y**, Hayes VM, Schuster SC, Miller WC (2010) Calling SNPs without a Reference Sequence. *BMC Bioinformatics*, 11:130
12. Miller WC, Wright S, **Zhang Y**, Schuster SC, Hayes VM. (2010) Optimization

- Methods for Selecting Founder Individuals for Captive Breeding or Reintroduction of Endangered Species. *Pac. Symp. Biocomput*, 15:43–53
13. Hsu CH, **Zhang Y**, Hardison R, Miller W (2009) Whole-Genome Analysis of Gene Conversion Events. *RECOMB-CG*, 181-192
  14. Cheng Y, Wu W, Kumar SA, Yu D, Deng W, Tripic T, King DC, Chen KB, Zhang Y, Drautz D, Giardine B, Schuster SC, Miller W, Chiaromonte F, **Zhang Y**, Blobel GA, Weiss MJ and Hardison RC (2009) Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications and mRNA expression. *Genome Res.* 19: 2172-2184.
  15. **Zhang Y**, Song GT, Vinar T, Green ED, Siepel A, Miller W (2009) Evolutionary History Reconstruction for Mammalian Complex Gene Clusters. *J Comp Biol*, 16:1051-1070.
  16. **Zhang Y**, Song GT, Hsu CH, Miller W (2009) Simultaneous History Reconstruction for Complex Gene Clusters in Multiple Species. *Pac. Symp. Biocomput* 14:162-173.
  17. **Zhang Y**, Niu T (2009) *Handbook on Analyzing Human Genetics Data – Computational Approaches and Software*. Chapter2: Haplotype Structure. Edited by Lin SL and Zhao HY, Springer.
  18. **Zhang Y** (2008) Poisson Approximation for Significance in Genome-wide ChIP-chip Tiling Arrays. *Bioinformatics*, 24(24):2825-2831.
  19. **Zhang Y** (2008) Tree-guided Bayesian inference of population structures. *Bioinformatics*, 24:965-971.
  20. **Zhang Y**, Song GT, Vinar T, Green ED, Siepel A, Miller W (2008) Reconstructing the evolutionary history of complex human gene clusters. *Proceedings of RECOMB 2008*, 29-49.
  21. **Zhang Y**, Liu J (2007) Bayesian inference of epistatic interactions in case-control studies. *Nat Genet.* 39:1167-1173.
  22. **Zhang Y**, Niu T, Liu J (2006) A coalescence-guided hierarchical Bayesian method for haplotype inference. *Am J Hum Genet*, 79(2):313-322.
  23. Valouev A\*, **Zhang Y\***, Schwartz DC, Waterman MS (2006) Refinement of optical map assemblies. *Bioinformatics*. 22:1217-1224. (\*equal contribution)
  24. Valouev A, Li L, Liu Y, Schwartz DC, Yang Y, **Zhang Y**, Waterman MC (2006) Alignment of optical maps. *J. Comp. Biol.* 13(2):442-462.
  25. **Zhang Y**, Waterman MS (2005) An Eulerian path approach to local multiple alignment for DNA sequences. *Proc Natl Acad Sci.* 102(5):1285-1290.
  26. **Zhang Y**, Waterman MS (2003) An Eulerian path approach to global multiple alignment for DNA sequences. *J. Comp. Biol.*, 10(6): 803-819.
  27. **Zhang Y**, Waterman MS (2003) DNA sequence assembly and multiple DNA sequence alignment by an Eulerian path approach. *Cold Spring Harb Symp Quant Biol.*, 68:205-212.
  28. Waterman MS, **Zhang Y** (2003) Eulerian path methods for multiple sequence alignment. *Proceedings of the 2003 IEEE Bioinformatics Conference, CSB 2003*, IEEE Computer Society, Order Number PR02000, Aug 11-14.

## **INVITED TALKS**

### 2012:

- 07/03-07/06 The 2<sup>nd</sup> Institute of Mathematical Statistics Asia Pacific Rim Meeting. Tsukuba, Japan.
- 03-29-04/01 Thirty Years of Computational Biology and Bioinformatics at USC: A symposium in honor of Professor Michael Waterman's 70<sup>th</sup> and Simon Tavare's 60<sup>th</sup> birthdays.

### 2011:

- 10/19 Department of Biostatistics, University of Minnesota, Twin Cities, MN.
- 10/10-10/15 The 12<sup>th</sup> International Congress of Human Genetics, the 61<sup>st</sup> Annual Meeting of the American Society of Human Genetics, Montreal, QC Canada.
- 09/30 Department of Statistics, Temple University, Philadelphia PA, USA.
- 08/29-09/01 The 7th International Conference on Multiple Comparison Procedures, Washington, DC, USA.
- 07/30-08/04 Joint Statistical Meeting 2011, Invited session, Miami, USA.
- 06/26-06/29 The 20th Annual ICOSA Applied Statistics Symposium. New York City, NY USA.
- 05/26 Department of Statistics, Peking University, Beijing, China.
- 02/28 Department of Statistics, Yale University, New Haven, CT USA.

### 2010:

- 09/26-09/29 European Conference on Computational Biology 2010, Ghent, Belgium.
- 07/10-07/12 International Conference on Statistics and Society, the 4<sup>th</sup> International Forum on Statistics, Renmin University, Beijing, China.
- 07/04-07/06 Probability and Statistics – an International Conference in Honor of PL Hsu's 100<sup>th</sup> Birthday. Peking University, Beijing, China.

### 2009:

- 06/28-07/01 The 1<sup>st</sup> Institute of Mathematical Statistics-Asia Pacific Rim Meeting, Seoul, Korea.
- 06/26-06/28 International Workshop on Probability Theory, Statistics and Their Application to Biology. Peking University, Beijing, China.
- 04/16-04/17 Current Topic Workshop: Systems Biology of Biological Processes and Diseases: Biological Problems and Statistical Solutions, Mathematical Biosciences Institute. Ohio State University, Ohio, OH USA.
- 01/06 The Pacific Symposium on Biocomputing (PSB) 2009, Big Island, Hawaii, USA.

### 2008:

- 10/16 Department of Statistics, University of Illinois Urbana Champaign. Urbana, IL USA.

10/10 Krishnaiah and Khatri Lecture Series: Frontiers of Evolutionary Biology. Department of Statistics, Penn State University, University Park, PA USA.

04/18 Center of Bioinformatics, George Washington University. Washington, DC, USA.

03/30 The 12<sup>th</sup> Annual International Conference on Research in Computational Molecular Biology (RECOMB) 2008, Singapore.

01/02 Department of Statistics, Peking University, Beijing, China.

2007:

02/07 Department of Biostatistics, Penn State University Hershey Medical School. Hershey, PA USA.

2006:

03/01 Center of Genomics, University of Wisconsin Madison, Madison, WI USA.

02/27 Department of Computer Science, University of Minnesota. Twin Cities, MN USA.

02/14 Department of Biostatistics, Johns Hopkins University. Baltimore, MD USA.

02/07 Department of Statistics, University of California Davis. Davis, CA USA.

01/26 Department of Biostatistics, Boston University. Boston, MA.

01/19 Department of Statistics, Penn State University. University Park, PA USA.

## **GRANTS**

### PENDING

NSF CAREER PI: Zhang Y 02/01/2012-01/31/2017  
 Effort: 25% of 12 months  
 Role: PI  
 Average Annual Support: \$103,458  
 Title: CAREER: Bayesian Graphical Variable Selection in Ultra-High Dimensional Data with Applications in Molecular Biology

### ACTIVE

NIH 1UL1RR033184-01 PI: Sinoway L 06/01/2011-02/29/2016  
 Effort: 25% of 12 months  
 Role: Investigator  
 Average Annual Support: \$4,467,207  
 Title: Penn State Clinical and Translational Science Institute

NIH RO1 DK065806-06 PI: Hardison R 02/01/2004-03/31/2014  
 Effort: 5% of 12 months  
 Multi-PI: Hardison RC, Schuster SC, Miller WC, Chiaromonte F, **Zhang Y**, Taylor J  
 Average Annual Support: \$706,720

Title: Global Predictions and Tests of Erythroid Regulation

NIH RO1 HG004718-01 PI: Zhang Y 08/15/2008 –06/30/2011  
Effort: 25% of 12 months (one year no cost extension)

Role: PI

Average Annual Support: \$141,595

Title: Bayesian Methods for Epistasis Association Mapping

NIH/NHGRI R01 HG02238-18S1 PI: Webb M 08/15/2008-06/30/2011  
Effort: 16.7% of 12 months (one year no cost extension)

Role: Co-PI

Average Annual Support: \$709,718

Title: Software for Analyzing Biosequence Data

#### COMPLETED

NIH RO1 1RC2HG0055573-01 09/26/2009-07/31/2011  
Multi-PI: Hardison RC, Schuster SC, Miller WC, Chiaromonte F, Zhang Y, Taylor J  
Title: ARRA: Enhance Human ENCODE by Function Comparisons to Mouse

#### **PROFESSIONAL SERVICES**

##### Conference and Journal Editorships:

- 2011 Session organizer and chair, the 20th annual ICSA applied statistics symposium, New York, USA, 06/26-06/29  
Program committee, the international conference on information science and technology, Nanjing, China, 03/26-03/28  
Editorial Board member of *PLoS One*  
Review Editor for *Frontiers in Genetics*  
Review Editor for *Frontiers in Statistical Genetics and Methodology*  
Review Editor for *Frontiers in Genetic Architecture*  
Associate Editor for *Journal of Biometrics & Biostatistics*
- 2010 Associate Editor for *Journal of Biometrics & Biostatistics*  
Program committee, CISP'10-BMEI, Yantai, China, 10/16-10/18
- 2009 Session chair, the international workshop on probability theory, statistics and their application to biology, Peking University, Beijing, China, 06/26-06/28  
Session chair, the 1<sup>st</sup> institute of mathematical statistics-Asia Pacific rim meeting, Souel, Korea, 06/28-07/01

##### Journal Refereeing (Selected)

American Journal of Human Genetics  
Annals of Applied Statistics  
Bioinformatics  
Biometrics  
Genetics

Genome Research  
Nucleic Acids Research  
Journal of American Statistical Association  
PLoS Genetics  
Proceedings of the National Academy of Sciences, USA

### **DEPARTMENT SERVICES**

2011	Department colloquium chair Graduate admission committee member Undergraduate committee member
2010	Graduate admission committee member Undergraduate committee member
2009	Department colloquium chair Undergraduate committee member
2008	Clifford C. Clogg memorial lecture series chair Graduate master exam committee chair Undergraduate committee member
2007	Department colloquium chair Graduate master exam committee member Undergraduate committee member
2006	Library committee chair

### **TEACHING**

#### The Pennsylvania State University

Fall 2011	Stat/Math415: Introduction to Mathematical Statistics
Spring 2011	Stat416: Stochastic Modeling
Fall 2010	Stat/Math415: Introduction to Mathematical Statistics
Spring 2010	Stat/CSE/BIOL597E: Bioinformatics II (graduate)
Fall 2009	Stat/Math415: Introduction to Mathematical Statistics
Spring 2009	Stat/CSE/BIOL598B: Bioinformatics II (graduate)
Fall 2008	Stat511: Regression Analysis and Modeling (graduate)
Spring 2008	Stat/Math415: Introduction to Mathematical Statistics
Fall 2007	Stat544: Categorical Data analysis I (graduate)
Spring 2007	Stat/Math415: Introduction to Mathematical Statistics
Fall 2006	Stat/Math414: Introduction to Probability

#### University of Southern California

Fall 2001	Math 108: Introductory College Mathematics
Summer 2001	Math 118x: Fundamental Principles of the Calculus

### **ADVISING AND MENTORING**

#### Postdoctoral Fellows (Current)

Yeonok Lee, 2010-present (Joint with Debashis Ghosh)

PhD Thesis Advisor (Current)

Jialin Xu, PhD in Statistics, expected 2012

Tyler Mayls, PhD in IBIOS, expected 2013

PhD Committee Member (Current)

Pannapa Changpetch, PhD in Statistics

Marcus Nunes, PhD in Statistics

Liye Zhang, PhD in IBIOS

Luyi Wo, PhD in IBIOS

Kuanbei Chen, PhD in Computer Science

Giltae Song, PhD in Computer Science

Jayaram Raghuram, PhD in Computer Science

PhD Committee Member (Past)

Zhenhan Zhang, PhD in IBIOS, graduated in 2011

Xianyun Mao, PhD in Statistics, graduated in 2010

Katerina Sinclair, PhD in Human Development and Family Study, graduated in 2010

Master Committee Member (Current)

Venkatesh Muktali, MS in IBIOS

Master Committee Member (Past)

Seyed Safa Eslanbolchi, MS in Energy and Mineral Engineering, graduated in 2010

Yueping Zhang, MS in Biology, graduated in 2007

Graduate Academic Advisor (Current)

Chandra, Saksham

Hao, Han