

数学与系统科学研究院学术报告

报告题目: Efficient Designs of Gene-Environment Interaction Studies: Implications of Hardy-Weinberg Equilibrium and Gene-Environment Independence

报告人: Guolian Kang

时间地点: 10:00-11:00 AM, Mar 10, Room 712, Siyuan Building

摘要:

Genetic and environmental risk factors play synergistic roles in the development of complex human phenotypes. Recently, many susceptible genetic variants have been discovered from genome-wide association studies, and fully capitalizing these exciting findings calls for investigations on whether these variants exercise the same effect in populations that are differentially exposed to environmental risk agents. Here, we assess the power of four two-phase case-control design strategies for assessing gene-environment (G-E) interaction or for assessing genetic variants or environment variables in the presence of gene-environment interactions. With a bi-allelic SNP and a binary E, we obtain close-form maximum likelihood estimates of both the main effect and interaction odds ratio parameters under the constraints of G-E independence and Hardy-Weinberg Equilibrium, and use the Wald statistic for all tests. For assessing G-E interactions, we conclude that a balanced selection of cases would lead to great power improvement, and that the stratified sampling of controls does not help power compared with the random sampling. In particular, supplementing external control data to an existing case-control sample does not help improve the power for testing G-E interactions, although it leads to improved power for assessing effects of G or E in the presence of G-E interactions.

报告人简介:

Guolian Kang

Department of Biostatistics and Epidemiology
University of Pennsylvania
208 Blockley Hall, 423 Guardian Drive
Philadelphia, PA 19104 USA

Office: (215) 746-3519
Fax: (215) 573-1050
Email: kangguo@mail.med.upenn.edu

Education

- **Ph.D. in System Modeling and Control**
Institute of Systems Science, Chinese Academy of Sciences, China July 2006
Dissertation: Association Studies on Complex Diseases by Case-Control Designs
Advisor: Ji-Feng Zhang, Ph.D.
- **M.S. in Applied Mathematics**
Qufu Normal University, China July 2003
Thesis: Oscillation and asymptotic analyses for nonlinear difference equations
- **B.S. in Mathematics**, Qufu Normal University, China July 2000

Continuing Education

- The 2009 international workshop on statistical genetics and methodology of twin and family studies: the advanced course, University of Colorado at Boulder, Boulder 03/2009
- FBAT/PBAT short course, Harvard University, Boston 06/2008
- Haplotype analysis of population & pedigree data in association studies, 05/2008
The University of Alabama at Birmingham (UAB), Birmingham
- Advanced gene mapping/linkage course, the Rockefeller University, New York 12/2007

Research Interests

- **Statistical Genetics/Genomics**—statistical methods for studying susceptibility genes for complex traits. Specific topics include: study designs, association analysis, multiple testing/false discovery rate, gene- or region-based association tests, gene-gene interactions, combining information from linkage and association, population stratification/substructure, admixture mapping, pathway analysis, copy number variation analysis, next generation sequencing data analysis, gene-environment interaction analysis
- **Genetical genomics/systems biology**— eQTL mapping and gene network inference by incorporating genetic marker information via information theoretic approaches
- **Statistics Modeling for Complex Data**—longitudinal data analysis, microarray data analysis, next-generation sequencing data analysis, machine learning and its application to biology and medicine
- **Application of Statistical Physics to Genetics/Traditional Chinese Medicine**—mainly apply the main idea of statistical physics to genetics and traditional Chinese medicine from a dynamic systems point of view

Research Experiences

- **Postdoctoral researcher**, University of Pennsylvania 12/2009 -
 1. Association analyses, imputation analyses, gene-environment interaction analyses for **cardiovascular disease** by using some popular softwares such as PLINK, HAPLOVIEW, EIGENSTRAT, ADMIXTURE, SNAP, MACH (Collaborate with Drs. Jinbo Chen and Daniel J. Rader)
 2. Association analyses, imputation analyses, gene-environment interaction analyses for **Hypertensive heart disease** using programs same as the first one (Collaborate with Drs. Jinbo Chen and Daniel L. Dries)
 3. Methodology developments for study designs, risk prediction model, pathway-based analysis, rare variant analysis, admixture mapping, case-control association analysis with related individuals, copy number variation analysis, next-generation sequencing data analysis, and gene-environment interaction analysis (Collaborate with Dr. Jinbo Chen)
 4. Methodology developments for inferring gene association networks and eQTL mapping in genetical genomics (Collaborate with Dr. Jinbo Chen)
- **Postdoctoral scholar**, University of Alabama at Birmingham 07/2007 - 12/2009
 1. Detecting complex epistasis among genes, multiple testing problems, gene- or region-based association tests, combining information from linkage and association and other important problems in GWAS for complex human diseases
 2. Structured association testing in genome-wide association studies of admixed populations, population stratification
 3. Analyses of four genome-wide real datasets including **rheumatoid arthritis** from Genetic Workshop Analysis 16 and **coronary artery disease, hypertension and rheumatoid arthritis** from the Wellcome Trust Case-Control Consortium
 4. Analyses of two genome-wide real datasets with population substructure including **left ventricular hypertrophy** on African American and Caucasian and one dataset of **prostate cancer** on European and African
- **Visiting scholar**, Michigan State University, USA 03/2006 - 06/2007
 1. Mathematical and statistical modeling and analysis for complex data
 2. Multi-stage analysis strategies and gene- or region-based designs for GWAS for complex human diseases
- **Research assistant**, Chinese Academy of Sciences, China 09/2003 - 02/2006
 1. Association studies of neurodevelopment-related candidate genes for a role in schizophrenia (Co-Investigator, National Natural Sciences Foundation of China)
 2. Mathematical studies on causative genes mapping with interaction effects on complex diseases, especially association studies, haplotype-based studies, single nucleotide polymorphisms (SNPs) data analysis, SNP synergistic block studies
 3. Mathematical studies on disease, medicine and symptoms in traditional Chinese medicine using entropy theory

Professional Activities

- **Memberships**

American Statistical Association, 2008-; American Society of Human Genetics, 2008-; International Genetic Epidemiology Society, 2008-

- **Invited reviewer (/times)**

BMC Genomics (/1), Atherosclerosis (/3), Behavior Genetics (/1), Computational Statistics and Data Analysis (/3), NeuroQuantology (/1), Journal of Medical Genetics and Genomics (/1), Journal of Public Health and Epidemiology (/3), Journal of Applied Statistics (/2), Statistical Applications in Genetics and Molecular Biology (/4)

- **Invited section chair**

The 2009 Joint statistical meetings, Washington DC (Chair) Aug. 2009

The 24th Chinese Control Conference, Guangzhou, China (Co-chair) Jul. 2005

- **Invited faculty of 1000**

Journal of Medical Genetics and Genomics since Oct. 2009

Journal of Public Health and Epidemiology since Oct. 2009

- **Judge** UAB Graduate Student Research Days 2008

- **Teaching experience** Tutor for Calvin Elam 2008

Grants

- 2008 Career Awards at the Scientific Interface, Burroughs Wellcome Fund (not funded) Total Amount: \$500,000. Role: Sole PI. Period: 5 years.

System-centric genome-wide association studies

- 2005-2007 National Natural Sciences Foundation of China No. 30400149

Total Amount: ¥260,000. Role: Co-PI. Period: 3 years.

Association studies of neurodevelopment-related candidate genes for a role in Schizophrenia

Awards

- 2010 Genetic Analysis Workshop 17 travel award, USA

- 2009 The Science Unbound Foundation Best Paper Award in Statistical Genetics Research, USA

- 2008 Travel fellowship from International Genetic Epidemiology Society, USA

- 2008 Spring 2008 Career Enhancement Award, The University of Alabama at Birmingham, USA

- 2003-2006 Research Assistant Fellowship, Chinese Academy of Science, China

- 1998-1999 Outstanding Undergraduate Scholarship, Qufu Normal University, China

Skills

- **Languages:** Matlab, SAS, R/Splus

- **Operating Systems:** Windows, Unix/Linux
- **Typesetting Systems:** Latex, Microsoft Word, Excel, Power Point, Dreamweaver
- **Genetic Software:** MS, fastPHASE, PLINK, ANCESTRYMAP, EIGENSTRAT, HAP-SAMPLE, MERLIN, FBAT/PBAT, Haploview, SNAP, MACH, STRUCTURE

Conferences and Workshops

- Nov. 2010 The 60th Annual meeting of American Society of Human Genetics, Washington DC, USA
- Oct. 2010 Genetic Analysis Workshop 17, Boston, USA
- Aug. 2009 Joint Statistical Meetings, Washington DC, USA
- Sep. 2008 International Genetic and Epidemiology Society, 17th Annual Meeting, St. Louis, Missouri, USA
- Sep. 2008 Genetic Analysis Workshop 16, St. Louis, Missouri, USA
- May 2007 Computational Genomics Workshop, Michigan State University, Michigan, USA
- Sep. 2006 Impact of Statistical Science in the 21st Century, Michigan State University, Michigan, USA
- Jul. 2005 The 24th Chinese Control Conference, Guangzhou, China
- Dec. 2004 New Strategies for Complex Diseases, Chinese National Human Genome Center, Beijing, China
- Dec. 2004 2004 Annual Report of the Center of Bioinformatics, Institute of Systems Science, Beijing, China
- Oct. 2004 The 2nd Institute on Statistical Genetics, Beijing, China

Oral Presentations

Invited presentations

- “Two-stage design: SNP discovery of mini-exome data by using extreme phenotype sequencing”, Genetic Analysis Workshop 17, Boston 10/13/2010
- “A generalized sequential Bonferroni procedure for genome-wide association studies”, Department of Biostatistics and Epidemiology, University of Pennsylvania 09/16/2009
- “A generalized sequential Bonferroni procedure for genome-wide association studies incorporating information of Hardy-Weinberg disequilibrium among cases”, 2009 Joint Statistical Meetings, Washington DC 08/05/2009
- “Mixed analysis for genomewide association studies”, Department of Biostatistics, University of Alabama at Birmingham 05/08/2007
- “Two-stage designs for genome-wide association studies”, Department of Epidemiology, Michigan State University 03/27/2007
- “Evolution analysis of life systems via entropy theory”, The 24th Chinese Control Conference, Guangzhou, China 07/15/2005

Journal club/research day presentations

- “Robust genome-wide scans with genetic model selection in two-phase analysis using case-control design”, Journal Club of Section on Statistical Genetics, University of Alabama at Birmingham 09/03/2009
- “A generalized sequential Bonferroni procedure for genome-wide association studies”, 2009 SOPH Research Day, University of Alabama at Birmingham 04/09/2009
- “A generalized sequential Bonferroni procedure for genome-wide association studies incorporating information of Hardy-Weinberg disequilibrium”, 2009 Postdoctoral Research Day, University of Alabama at Birmingham 02/05/2009
- “Genome-wide weighted hypothesis testing in family-based association studies, with an application to a 100K scan”, Journal Club of Section on Statistical Genetics, University of Alabama at Birmingham 06/05/2008
- “A sequentially weighted multiple testing procedure and its application to genome-wide association studies”, 2008 Postdoctoral Research Day, University of Alabama at Birmingham 02/05/2008

Bibliographic Listings

- Selected Biography, Marquis Who’s Who in America, 2009

Selected Posters

1. **Guolian Kang**, Guimin Gao, Sanjay Shete, David T. Redden, Bao-Li Chang, Timothy R. Rebbeck, Jill S. Barnholtz-Sloan, Nicholas Patterson, Nicholas M. Pajewski, David B. Allison. (2010) Capitalizing on admixture in genome-wide association studies: A two-stage testing procedure and application to height in African-Americans. The 19th Annual Meeting of The American Society of Human Genetics. (Washington, DC)
2. Bo Jiang, Xiao Zhang, Yijun Zuo, **Guolian Kang**. (2010) A truncated tail strength method for testing an overall hypothesis in a dataset. The 19th Annual Meeting of The American Society of Human Genetics. (Washington, DC)
3. **Guolian Kang**, Guimin Gao (2009). A Bonferroni procedure using weights accounting for Hardy-Weinberg disequilibrium information in genome-wide association studies. The 18th Annual Meeting of The American Society of Human Genetics. (Honolulu, Hawaii)
4. **Guolian Kang**, Bo Jiang (2009). Analysis of two gene-centric approaches for genome-wide association studies. The 18th Annual International Genetic Epidemiology Society Meeting. (Kahuku, Hawaii)
5. Guimin Gao, **Guolian Kang** (2009) A generalized sequential Bonferroni procedure for genome-wide association studies incorporating information on Hardy-Weinberg disequilibrium among cases. The 18th Annual International Genetic Epidemiology Society Meeting. (Kahuku Hawaii)
6. **Guolian Kang** (2008). An efficient multilocus Monte Carlo approach for gene-centric genome-wide association studies. The 17th Annual International Genetic Epidemiology Society Conference. (St. Louis, Missouri)
7. Guimin Gao, **Guolian Kang** (2008). A generalized sequential Šidák procedure for multiple hypothesis testing. The 17th Annual International Genetic Epidemiology Society Conference. (St. Louis, Missouri)

8. Yuehua Cui, **Guolian Kang**, et al. (2007). An entropy approach for gene-centric genomewide association study. Cambridge Healthtech Institute 7th annual Conference for New Applications for Microarray Data Analysis: Integrating Genetics with Omics. (Washington DC)

Peer-Reviewed Publications

Abstracts

1. **Guolian Kang**, Guimin Gao, Sanjay Shete, David T. Redden, Bao-Li Chang, Timothy R. Rebbeck, Jill S. Barnholtz-Sloan, Nicholas Patterson, Nicholas M. Pajewski, David B. Allison. (2010) Capitalizing on admixture in genome-wide association studies: A two-stage testing procedure and application to height in African-Americans. The American Society of Human Genetics.
2. Bo Jiang, Xiao Zhang, Yijun Zuo, **Guolian Kang**. (2010) A truncated tail strength method for testing an overall hypothesis in a dataset. The American Society of Human Genetics.
3. **Guolian Kang**, Guimin Gao (2009). A Bonferroni procedure using weights accounting for Hardy-Weinberg disequilibrium information in genome-wide association studies. *American Journal of Human Genetics*. S82: 475.
4. **Guolian Kang**, Bo Jiang (2009). Analysis of two gene-centric approaches for genome-wide association studies. *Genetic Epidemiology* 32 : 795-796.
5. Guimin Gao, **Guolian Kang** (2009) A generalized sequential Bonferroni procedure for genome-wide association studies incorporating information on Hardy-Weinberg disequilibrium among cases. *Genetic Epidemiology* 32 : 790-790.
6. **Guolian Kang** (2008). An efficient multilocus Monte Carlo approach for gene-centric genome-wide association studies. *Genetic Epidemiology* 32 : 698-698,
7. Guimin Gao, **Guolian Kang** (2008). A generalized sequential Šidák procedure for multiple hypothesis testing. *Genetic Epidemiology* 32 : 690-690.

Original papers (Co-first author*, Corresponding author[¶])

1. Jinbo Chen*, **Guolian Kang***, Tyler VanderWeele, Cuilin Zhang, Bhramar Mukherjee. (2011). Gene-Environment independence and Hardy-Weinberg disequilibrium: Implications for the two-phase Case-Control design. *Statistics in Medicine* (Accepted)
2. Bo Jiang, Xiao Zhang, Yijun Zuo, **Guolian Kang**[¶]. (2010). A powerful truncated tail strength method for testing multiple hypotheses in a dataset. *Journal of Theoretical Biology* (In press)
3. **Guolian Kang**, Guimin Gao, Sanjay Shete, David T. Redden, Bao-Li Chang, Timothy R. Rebbeck, Jill S. Barnholtz-Sloan, Nicholas Patterson, Nicholas M. Pajewski, David B. Allison. (2010). Capitalizing on admixture in genome-wide association studies: A two-stage testing procedure and application to height in African-Americans. *Frontiers in Statistical Genetics and Methodology* (Accepted)
4. Yijun Zuo, **Guolian Kang**[¶]. (2010). A mixed two-stage analysis for detecting interacting genes in genomewide association studies. *Journal of Theoretical Biology* 262: 576-583.
5. **Guolian Kang**, Douglas K. Childers, Nianjun Liu, Kui Zhang, Guimin Gao. (2009) Genome-wide association studies of Rheumatoid Arthritis data via multiple hypothesis testing methods for correlated tests. *BMC Proceedings* 3: S38.

6. Douglas K. Childers, **Guolian Kang**, Nianjun Liu, Guimin Gao, Kui Zhang. (2009) Application of imputation methods for the analysis data from genome-wide association studies. *BMC Proceedings* 3: S24.
 7. Lisa J. Martin, Guimin Gao, **Guolian Kang**, Yixin Fang, Jessica G. Woo. (2009) Improving the signal-to-noise ratio in genomewide association studies. *Genetic Epidemiology* 33: S29 - S32.
 8. **Guolian Kang**, Bo Jiang, Guimin Gao. (2009) A generalized sequential Bonferroni procedure incorporating information of Hardy-Weinberg disequilibrium among cases for genome-wide association studies. In *JSM Proceedings*, Section on Statistics in Epidemiology. Washington, DC: American Statistical Association. 4434-4448.
 9. **Guolian Kang**, Keying Ye, David B. Allison, Nianjun Liu, and Guimin Gao. (2009) Weighted multiple hypothesis testing procedures. *Statistical Applications in Genetics and Molecular Biology*. 8(1), Article 23.†‡
- †Most Popular Papers (4/16/2009-8/21/2009) in the journal based on the average number of full text downloads per day since posted <http://www.bepress.com/sagmb/topdownloads.html>

‡The 10 Most Popular Articles in Collection of Biostatistics Research Archive (COBRA) on May 27, 2009.
10. **Guolian Kang**, Weihua Yue, Jifeng Zhang, Yuehua Cui, Yijun Zuo and Dai Zhang. (2008). An entropy-based approach for testing genetic epistasis underlying complex diseases. *Journal of Theoretical Biology* 250: 362-374.†
- †Awarded 2009 best paper in the area of statistical genetics from Science Unbound Foundation <http://www.scienceunboundfoundation.org/2009.html>
11. Yuehua Cui, **Guolian Kang**, Kelian Sun, Minping Qian, Roberto Romero and Wenjiang Fu. (2008). Gene-centric genome-wide association study via entropy. *Genetics* 179: 637-650. †
- †Research highlighted by *Nature Review Genetics* in 2008: 9. <http://www.nature.com/nrg/journal/v9/n6/full/nrg2387.html>
12. **Guolian Kang**, Weihua Yue, Jifeng Zhang, Huebner Marianne, Handi Zhang, Yan Ruan, Tianlan Lu, Yansu Ling, Yijun Zuo, Dai Zhang. (2008). Two-stage designs to identify the effects of SNP combinations on complex diseases. *Journal of Human Genetics* 53: 739-746.
 13. **Guolian Kang**, Shao Li, Jifeng Zhang. (2008). Entropy-based models for interpreting life systems in traditional Chinese medicine. *Evidence-based Complementary and Alternative Medicine* 5(3): 273-279. †‡
- †Reported by NewsRX.com <http://www.newsrx.com/article.php?articleID=1304851> in 10/2008.

‡Reported by international media “International Adaptogens” published by CMA (Adaptogens MEDICAL CENTER) <http://www.adaptogeno.com/svms/noticias/noticia367.asp> in 10/2008.
14. **Guolian Kang**, Yijun Zuo. (2007). Entropy-based joint analysis for two-stage genome-wide association studies. *Journal of Human Genetics* 52: 747-756.
 15. Weihua Yue*, **Guolian Kang***, Yanbo Zhang, Mei Qu, Fulei Tang, Yonghua Han, Yan Ruan, Tianlan Lu, Jifeng Zhang and Dai Zhang. (2007). Association of DAOA polymorphisms with schizophrenia and clinical symptoms or therapeutic effects. *Neurosci Lett* 416: 96-100.
 16. Weihua Yue*, **Guolian Kang***, Handi Zhang, Fulei Tang, Mei Qu, Yonghua Han, Jun Yan, Yan Ruan, Tianlan Lu and Dai Zhang. (2007) Association study between

- schizophrenia and NRG1,G72,RGS4 polymorphisms. *Chin J of Behavioral Med Sci.* 16: 418-820. (In Chinese)
17. Weihua Yue, Zhonghua Liu, **Guolian Kang**, Jun Yan, Fulei Tang, Yan Ruan, Jifeng Zhang, Dai Zhang. (2006). Association of G72/G30 polymorphisms with early-onset and male schizophrenia. *NeuroReport* 17(18): 1899-1902.
 18. **Guolian Kang**, Fanwei Meng. (2006). Oscillation criteria for second-order quasi-linear neutral difference equations. *Journal of Engineering Mathematics* 23(1): 923-926.
 19. **Guolian Kang**. (2006). Oscillation criteria for second-order half-linear neutral difference equations. *Journal of Mathematical Research and Exposition* 26(2): 247-252.
 20. **Guolian Kang**. (2005). Oscillation criteria for second-order nonlinear difference equation with “summation small” coefficient. *Bull. Korean Math. Society* 42(2): 245-256.
 21. **Guolian Kang**, Shao Li, Jifeng Zhang. (2005). Evolution analysis of life systems via entropy theory. *Proc. of the 24th Chinese Control Conference* 1567-1572.
 22. **Guolian Kang**, Hui Zhang. (2004). Oscillation criteria of solutions of nonlinear difference equations of second order. *Annals of Differential Equations* 20(1): 41-48.
 23. **Guolian Kang**. (2004). Oscillation theorems of solutions of nonlinear difference equations of second order. *Journal Of Binzhou Teachers College* 18(4): 6-10.
 24. **Guolian Kang**. (2004). Some generalizations and applications on some integral inequalities. *Journal of Engineering Mathematics* 21(5): 715-721. (In Chinese)

Papers in Preparation (Co-first author*, Corresponding author[¶])

1. **Guolian Kang**, Bo Jiang, Jinbo Chen. (2011). A powerful information-theoretic metric for testing gene-environment interactions via Hardy-Weinberg Equilibrium and Gene-Environment Independence. (In preparation)
2. **Guolian Kang**, Bo Jiang, Jinbo Chen. (2011). A system information metric for inferring networks and genetic association in genetical genomics. (In preparation)
3. John Stylianou, **Guolian Kang**, et al. (2010). Dense SNP genotyping of Cholesteryl Ester Transfer Protein. (In preparation)
4. Jinbo Chen, **Guolian Kang**, et al. (2010). Genetic Ancestry: is it Useful for Risk Stratification? (In preparation)
5. **Guolian Kang**[¶], Dongyu Lin, Mingyao Li, Yubin Yao, Jinbo Chen. (2010). Two-stage design: SNP discovery of mini-exome data by using extreme phenotype sequencing. *BMC Proceeding* (In review)
6. **Guolian Kang**, Jinbo Chen, Daniel Dries, et al. (2011). Admixture mapping of African Americans identifies Left Ventricular Hypertrophy susceptibility loci on chromosomes 6 and 8. (In preparation)
7. **Guolian Kang**, Bo Jiang, Yuehua Cui. (2010). Gene-centric association analysis: a comparison study. (Submitted)
8. **Guolian Kang**, Bo Jiang, Nianjun Liu, David B Allison, Guimin Gao. (2011). A generalized sequential Bonferroni procedure incorporating information of Hardy-Weinberg disequilibrium among cases for genome-wide association studies. (To be submitted)

References

- Professor Jinbo Chen
Department of Biostatistics and Epidemiology
School of Medicine
University of Pennsylvania
Philadelphia, PA 19104
Phone: (215) 746-3915
Email: jinboche@mail.med.upenn.edu
- Professor Daniel L. Dries
Hospital of the University of Pennsylvania
Heart Failure/Transplant Center
3400 Spruce Street
Philadelphia, PA 19104-4283
Phone: (215) 662-2727
Email: Daniel.Dries@uphs.upenn.edu
- Professor Guimin Gao
Department of Biostatistics
Virginia Commonwealth University
Richmond, Virginia 23284
Phone: (804) 827-2775
Email: ggao3@vcu.edu
- Professor Daniel J. Rader
Institute for Translational Medicine and Therapeutics
421 Curie Boulevard
University of Pennsylvania
Philadelphia, PA 19104
Office: (215) 573-4176
Email: rader@mail.med.upenn.edu
- Professor Jifeng Zhang
Institute of Systems Science
Chinese Academy of Science
Beijing, 100080 China.
Tel: 86-10-62565782
Email: jif@iss.ac.cn
- Professor Yijun Zuo
Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824
Phone: (517) 432-5413
Email: zuo@msu.edu