

Department of Statistics, 317 Evans Hall
University of California, Berkeley
Berkeley, CA 94720

Phone (510) 642-6433
E-mail hhuang@stat.berkeley.edu
<http://www.stat.berkeley.edu/~hhuang>

Haiyan Huang

Current Position

Assistant Professor (07/2003-present)

Department of Statistics
Interdepartmental Group in Biostatistics
Graduate Group in Computational and Genomic Biology
University of California, Berkeley, USA

Education/ Training

9/2001-7/2003 Harvard University
Post-Doctoral Research Fellow, Departments of Stat and Biostat (computational biology)
Supervisors: Wing H. Wong, Jun S. Liu

9/2000-12/2000 Institute for Pure and Applied Mathematics, UCLA
Research Fellow in Functional Genomics

8/1997-8/2001 University of Southern California
PhD, Applied Mathematics (computational biology)
Advisor: Larry Goldstein
Dissertation: Bounds for the errors in word count distributional approximations.

8/1993-7/1997 Peking University, China
BS, Mathematics

Awards & Grants Received

- **NIH R21EY019094 (PI)**, \$421K, 4/2009-3/2011
"Applied Statistics to a Secondary Analysis of Public Repositories for Microarray Data"
- **EBI-UK (NIH Prime; ENCODE DAC) (Investigator)**, \$629K, 2008-2012
"Develop Statistical and Computational Methodologies for Analyzing ENCODE Data"
- **NSF BECR11652 (co-PI)**, \$ 210K, 2007-2010
"Collaborative Research: NanoCytometry: A Point-of-Care Technology for Monitoring Chronic Leukemia Patients"
- **USC (NIH Prime; 1R01GM074163) (subcontract PI)**, \$68K, 1/2006-12/2010
"Integrative Analysis of Cross-Platform Microarray Data"
- **NIH 5R01GM075312 (co-PI)**, \$671K, 6/2005-5/2009
"Comparative Genomics to Identify Functional Blocks and Horizontal Gene Transfer"
- **Junior Faculty Research Award**, University of California, Berkeley, 2004
"Developing statistical methods for SAGE data"
- **Presidential Chair Fellow**, University of California, Berkeley, 2003-2004
- **Fellowship** from Institute for Pure and Applied Mathematics.
Functional Genomics Program, Fall 2000
- **Meritorious Award** in the US Mathematical Contest in Modeling, 1997
- **Meritorious Award** in the Chinese Mathematical Contest in Modeling, 1996

Research Interests

Development of novel statistical and computational methods to genome biology. Study of mathematical and statistical theories that underlie those methods.

Specific bioinformatics topics of current interest (addressed via integrative analysis of various types of genomic data) include:

- Prediction of functional elements in genome sequences;
- Identification of gene relationships, pathways and networks;
- Translational bioinformatics: analysis of multiple-microarrays and integrative information retrieval for aiding disease diagnosis.

I am also interested in analyzing next-generation high-throughput sequencing data and other interesting data generated from new bio-technologies.

Publications

1. Bickel P, Brown B, Huang H, Li Q (2009). An Overview of Recent Developments in Genomics and Associated Statistical Methods. **Philosophical Transactions of the Royal Society A**. In press.
*authors ordered alphabetically
2. Teng S, Huang H* (2009). A Statistical Framework to Infer Functional Gene Associations from Multiple Biologically Interrelated Microarray Experiments. **Journal of the American Statistical Association (JASA)**, June 2009, Vol. 104, No. 486.
*corresponding author
3. Wang F, Jiang T, Sun Z, Teng SL, Luo X, Zhu Z, Zang Y, Zhang H, Yue W, Hong N, Huang H, Blumberg H, Zhang, D (2009). Neuregulin 1 Genetic Variation and anterior cingulum integrity in patients with schizophrenia and healthy controls. **Journal of Psychiatry & Neuroscience**, 2009 May;34(3):181-6.
4. Liu C, Hu J, Kalakrishnan M, Huang H*, Zhou XJ* (2009) Integrative Disease Classification Based on Cross-platform Microarray Data. **BMC Bioinformatics**, 2009 Jan;10 Suppl 1:S25.
*co-corresponding authors
5. Carbonaro A, Mohanty SK, Huang H, Godley LA and Sohn LL (2008) Cell Characterization Using A Protein-Functionalized Pore, **Lab Chip**, 8(9):1478-85.
6. Huang H, Cai L, Wong WH. (2008) Clustering analysis of SAGE transcription profiles using a Poisson approach. **Methods Mol Biol**. 2008; 387:185-98. (Book Chapter)
7. Huang Y, Li H, Hu H, Yan X, Waterman MS, Huang H, Zhou XJ (2007). Systematic Discovery of Functional Modules and Context-Specific Functional Annotation of Human Genome. **Bioinformatics**, 23(13):i222-i229.
8. ENCODE Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature**. 447, 799-816.
9. Kim K, Zhang S, Jiang K, Cai L, Lee IB, Feldman LJ, Huang H* (2007). An Efficient Measure of Similarity between Gene Expression Profiles through Data Transformations. **BMC Bioinformatics**, 8:29 (highly accessed paper).
*corresponding author
10. Jiang K, Zhang S, Lee S, Tsai G, Kim K, Huang H, Zhu T, Feldman LJ (2006). Transcription Profile Analyses Identify Genes and Pathways Central to Root Cap Functions in Maize. **Plant Molecular Biology**, 60(3):343-63.

11. Huang H, Kim K (2006). Unsupervised clustering analysis of gene expression, **Chance**, vol. 19, No.3.
12. Zhou XJ, Kao MJ, Huang H, Wong A, Nunez-Iglesias J, Aparicio OM, Morgan TE, Wong WH (2005). Functional annotation and network reconstruction through cross-platform integration of microarray data. **Nature Biotechnology**, 23(2):238-43.
13. Zhao X, Huang H, Speed T (2005). Finding short DNA motifs using permuted Markov models. **Journal of Computational Biology**, 12(6): 894-906 (journal version of the 2004 RECOMB paper; number 16 below)
14. Huang H, Kao MJ, Zhou X, Liu JS, Wong WH (2004). Determination of local statistical significance of patterns in Markov sequences with application to promoter element identification." **Journal of Computational Biology**, 11(1):1-14.
15. Cai L*, Huang H*, Blackshaw S, Liu JS, Cepko CL, Wong WH (2004). Clustering analysis of SAGE data using a Poisson approach. **Genome Biology**, 5(7):R51
*Joint first authors
16. Zhao X, Huang H, Speed T (2004). Finding short DNA motifs using permuted Markov models. **Proceedings of RECOMB 2004**.
17. Blackshaw S, Harpavat S, Trimarchi J, Cai L, Huang H, Kuo W, Fraioli R, Cho S, Yung R, Asch E, Wong WH, Cepko CL (2004). Genomic analysis of mouse retinal development. **PLoS Biol**, 2(9):E247.
18. Allinen M, Beroukhi R, Cai L, Brennan C, Domenici CJ, Huang H, Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers W, Polyak K (2004). Molecular characterization of the tumor microenvironment in breast cancer. **Cancer Cell**, 6(1):17-32.
19. Lippert RA, Huang H, Waterman MS (2002). Distributional regimes for the number k -word matches between two random sequences. **Proc Natl Acad Sci.USA**, 99(22):13980-9.
20. Huang H (2002). Error bounds on multivariate normal approximations for word count statistics. **Advances in Applied Probability**, 34(3): 559-586.

Books Edited

- "Research in Computational Molecular Biology" (11th Annual International Conference, RECOMB 2007), edited by Terry Speed and Haiyan Huang, Published by Springer.

Articles Submitted & Manuscripts in Preparation

- Bickel PJ, Boley N, Brown JB, Huang H, Zhang NR (2009). Non Parametric Methods for Genomic Inference.
*authors ordered alphabetically
- Huang H, Liu C, Zhou XJ (2009). Transforming public microarray repositories into disease diagnosis databases.
- Jiang K, Zhu T, Huang H, Feldman L (2008). The Maize Root Stem Cell Niche: A partnership between two transcriptionally distinct stem cell Populations.
- Na Xu, Bickel PJ, Huang H. Semi-parametric method for multiple RNA tiling arrays.
- Brown JB, Kechris KJ, Poulin F, Xu N, Bickel PJ, Huang H. Classification of Transcriptional Function in Human Conserved Non-Coding Elements.

Software Developed

GEA: Gene Expression Analyzer; for clustering/classification/significance analysis of SAGE and Microarray gene expression data (<http://cell.rutgers.edu/gea>)

Knorm: An appealing statistical method for gene association inference across multiple dependent experimental conditions.
(<http://cran.r-project.org/web/packages/knorm>)

LMM: For predicting transcription factor binding sites by evaluating a candidate site in a local genomic context. (Available upon request)

TilingAnalyzer: For transcriptome detection by analyzing multiple RNA tiling arrays
(Available upon request)

Professional Activities

- **Editor** for “Statistical Applications in Genetics and Molecular Biology”, 2004-present. (A popular journal in Statistical and Computational Biology)
- **Guest editor** of Special Issue on “Information Integration and Meta-analysis of Multiple High-throughput Biological Experiments”, *Advances in Bioinformatics*.
- **Organizer of the invited session** on “Integrative Analysis of Various Types of Genomic Data,” ICSA conference, December 2010, Guang Zhou, China.
- **Local Committee Chair**, ICSA 2009 Applied Statistics Symposium, June 21-24, San Francisco, California. (led a committee of 6 faculty from Stanford U, UC Berkeley, UC Davis and Cal State U to manage the local activities; there were over 400 participants)
- **IMS Program Chair** for 2009 WNAR/IMS annual meeting, Portland State University, June 14-17, 2009. (A prestigious national conference in Statistics; I designed the IMS program that includes 5 invited sessions, 2 joint IMS-WNAR invited sessions, and the esteemed Medallion Lecture.)
- **Organizer of the invited session** on “Statistics for Genomics,” The seventh World Congress in Probability and Statistics, Singapore, July 14-19, 2008. (A prestigious international conference in Probability and Statistics)
- **Participant** in the NSF/CAUSE-sponsored workshop on “Integrating Computing in the Statistics Curricula,” Berkeley, CA, July 13-17, 2008.
- **Organizer of the invited session** on “Expression and Sequence Analysis,” The WNAR/IMS annual meeting, Davis, California, June 22-25, 2008.
- **Organization Committee Member, co-Editor of the proceedings, Host of a keynote speech**, RECOMB 2007. (One of the most popular and prestigious international conferences in Computational Biology; over 700 participants)
- **NIH volunteer**, June 2006.
- **Program Committee member** for the Fourth Asia-Pacific Bioinformatics Conference (APBC'06), Taipei, Taiwan, Feb. 13-16, 2006.
- **IMS Organization Committee member**, joint meeting of CSPS/IMS, July 9-12, 2005.
- **Reviewer** for *Journal of the American Statistical Association*; *Advances in Applied Probability*; *Annals of Applied Probability*; *Journal of Applied Probability*; *Journal of the Royal Statistical Society (Series C)*; *Proceedings of National Academy of Sciences*; *Biometrics*; *Nature Biotechnology*; *Genome Research*; *Genome Biology*; *RECOMB*; *Bioinformatics*; *Nucleic Acid Research*; *BMC Bioinformatics*; *Journal of Computational Biology* *IEEE journals*; *Biotechniques*; *Computational Statistics and Data Analysis*; *Statistics in Medicine*; *ICARCV (conference) 2004*; *Statistical Applications in Genetics and Molecular Biology*; *Journal of computer science and Technology*; *NSF proposals*; *Springer books*; etc.

Department and Campus-wide Academic Services

Course Committee Member, Department of Statistics, Fall 2009

Co-designed a new course "Statistical Genomics". (This course is the first in a two-semester sequence and provides an introduction to statistical and computational methods in classical genetics/genomics and population genetics.)

Executive Committee Member, Designated Emphasis in Computational Biology and the center for Computational Biology, UC Berkeley & QB3, 2008-2009

Retreat organizer, Designated Emphasis in Computational Biology and the center for Computational Biology, UC Berkeley & QB3, Spring 2007

Meeting with prospective students for the Berkeley Edge Program, Spring 2007

VIGRE committee Member, Department of Statistics, Fall 2006

Diversity Officer, Department of Statistics, Fall 2006

Student Admission Committee Chair, Department of Statistics, 2006-2007

Member of the Hiring Committee for the Junior Faculty Position in Biostatistics, 2005-2006

Neyman Seminar Organizer, Department of Statistics, UC Berkeley, Spring 2005

Student Admission Committee Member, Department of Statistics, 2004-2006

Organizer of Statistics women faculty/students lunch meeting, Spring 2004

Colloquium Organizer, Department of Statistics, UC Berkeley, 2003-2005, Spring 2008, 2009

Teaching & Advising

Undergraduate courses

STAT 152: Survey Sampling, Fall 2003-2006

BIOE/STAT C141: Statistics for Bioinformatics, Spring 2004-2008

STAT 131A: Statistical Inferences for Social and Life Scientists, Spring 2009

STAT 157: Seminar on Topics in Probability and Statistics, Fall 2009

Master courses

STAT 200B: Introduction to Probability and Statistics at an Advanced Level, Spring 2006, 2007

PhD courses

STAT 210A: Theoretical Statistics, Fall 2008, 2009

STAT 246: Statistical Genetics, Spring 2009

Former Students

PhD Students

Siew-Leng Melinda Teng (PhD 2007)

Thesis: "Statistical methods in integrative analysis of gene expression data with applications to biological pathways"

Current position: statistician at Genentech Inc.

Na Xu (PhD 2008; co-advised with Prof. Peter Bickel)

Thesis: "Transcriptome Detection by Multiple RNA Tiling Array Analysis and Identifying Functional Conserved Non-coding Elements"

Current position: statistician at Genentech Inc.

Master Students

Hua Chen (Master 2008)

Thesis: "A Bayesian Method for Multi-Loci Association Study of Human Disease"

Current position: research fellow at Harvard University

Undergraduate Students

Dan Nguyen (Undergraduate 2007), RA, summer 2007

Louis Ho (Undergraduate, supported by VIGRE program) Spring 2006

Denis Lanki (Undergraduate, supported by VIGRE program) Fall 2006

Phillip Rescober (Undergraduate 2004)

Honor thesis: "An analysis of belief-updating in discrimination issues"

Visitors

Kyungpil Kim (Visiting graduate student, POSTECH, Korea, Jan 2005 – Dec 2007)

Jing Zhou (Visiting graduate student, Shang Hai Jiao Tong University, China,
Oct 2007 – Sep 2008; co-advised with Prof. Peter Bickel)

Current Students and Post-docs

Graduate Students

Daisy Yan Huang (Statistics PhD program, UC Berkeley, Mar 2008 – present)

Kyungpil Kim (Biostatistics PhD program, UC Berkeley, Aug 2008 – present)

Ling Meng (Statistics Master program, UC Berkeley, Dec 2008 – present; she is also a
PhD candidate in Plant and Microbial Biology, UC Berkeley)

Qinghui Gao (Visiting graduate student, Beijing University of Aeronautics and
Astronautics, Nov 2008 – Oct 2009)

Post-docs

Qunhua Li (Sep 2008 – present; co-advised with Prof. Bickel)

Ci-Ren Jiang (appointment starts in September, 2009)

Major Collaborators

Peter Bickel research group (Statistics, UC Berkeley)

X. Jasmine Zhou research group (Molecular & Computational Biology, USC)

Lewis Feldman research group (Plant & Microbial Biology, UC Berkeley)

Li Cai research group (Biomedical Engineering, Rutgers University)

Lydia Sohn research group (Mechanical Engineering, UC Berkeley)

Serving as Dissertation/Thesis Committee Members

Xin Victoria Wang, PhD in Biostatistics, expected in Fall 2009

Soyeon Ahn, PhD in Statistics, May 2009

Hui Tang, PhD in Biostatistics, Dec 2008

Joseph Mark Dale, Master in Bioengineering, May 2008

Min Liu, PhD in Civil Engineering, May 2007

Ariel Evnine, Master in Statistics, Dec 2006

Serving as Oral Examiner for Doctoral Candidates

Yue Wang (Bioengineering, qual. Exam) Aug 2009

Jingyi Jessica Li (Biostatistics, qual. Exam, chair) Aug 2009

ChuiHsin Chen (Civil Engineering, qual. Exam, outside member) Aug 2009

Sangjoon Park (Civil Engineering, qual. Exam, outside member) Aug 2009

Johann GagnonBartsch (Statistics, qual. Exam) May 2009

Oleg Sergeyevich Mayba (Statistics, qual. Exam) May 2009

Karl Rohe (Statistics, qual. Exam) Feb 2009

Nancy Feng Dan Dong (Civil Engineering, qual. Exam, outside member) August 2008

Nancy Wang (Statistics, qual. Exam) May 2008
 Frances Tong (Statistics, qual. Exam, chair) April 2008
 Xin Victoria Wang (Statistics, qual. Exam) April 2008
 James Bentley Brown (Applied Science and Technology, qual. Exam, outside member) October 2007
 David Gregory Purdy (Statistics, qual. Exam) September 2007
 Guoyuan Wu (Mechanical Engineering, qual. Exam, outside member) August 2007
 Qixing Zheng (Mechanical Engineering, qual. Exam, outside member) April 2007
 Naraporn Somboonna (Bioengineering, qual. Exam, served as outside member) April 2007
 Margaret Taub (Statistics, qual. exam, committee chair), January 2007
 Wassim Michael Ghannoum (Engineering-Civil and Environmental Engineering, qual. Exam, outside member) June 2006
 Zile Wei (Computer Science, qual. Exam, outside member) March 2006
 Sandra Sinisi (Biostatistics, qual. exam), December 2005
 Soyeon Ahn (Statistics, qual. exam, committee chair), November 2005
 Guillaume Obozinski (Statistics, qual. exam), August 2005
 Hui Tang (Biostatistics, qual. exam, committee chair), August 2005
 Min Liu (Civil Engineering, qual. Exam, outside member), May 2005
 Na Xu (Statistics, qual. exam, committee chair), August 2004
 Annette Molinaro (Biostatistics, qual. Exam), December 2003;
 Current position: faculty at Yale University

Invited talks

1. Workshop on "Mathematical Genomics", Mathematical Sciences Research Institute, April 13 – 15, 2009.
2. Department Seminar, Biostatistics, UCSF, Oct 2008
3. WJAR/IMS Joint Annual Meeting, Davis, June 2008
4. Department Seminar, Statistics, UC Davis, June 2008
5. Newton Institute Workshop on "High Dimensional Statistics in Biology", Cambridge, UK, March 31- April 4, 2008
6. Department Seminar, Statistics, University of Illinois at Urban- Champaign, Oct 2007
7. Keynote speaker for the workshop/conference on "Statistics for Biomolecular Data Integration and Modeling," Monte Verita, Ascona, Switzerland, June 2007
8. Department Seminar, Center of Statistical Sciences , Brown University, April 2007
9. Research Seminar, GENETECH, January 2007
10. Biostatistics Seminar, UCSF, November 2006
11. Department Seminar, Statistics, Stanford University, October 2006
12. ICSA 2006 Applied Statistics Symposium, University of Connecticut, June 14-17, 2006
13. Workshop on Frontiers of Statistics, Princeton University, May 18-20, 2006
14. Workshop on "Sequence Analysis Toward System Biology", IPAM, Jan 2006
15. Statistics Department Seminar, UCLA, Oct 2005
16. Department Seminar, Computational Molecular Biology and Bioinformatics, USC, Oct 2005
17. Statistics Department Seminar, Duke University, Sept 2005
18. Joint Statistical Meeting, Minnesota, Aug 2005
19. IMS/CSPS Joint Meeting, Beijing, China, July 2005
20. WJAR/IMS Joint Meeting, Alaska, June 2005
21. Biostatistics Seminar, SCHARP/FHCRC, Seattle, May 2005

22. Department Seminar, Statistics, University of California at Davis, May 2005
23. Department Seminar, Biostatistics, Stanford University, November 2004
24. Department of Biostatistics, University of Pittsburgh. October 2004
25. SAGE 2004 conference on decoding the "OMEs", Boston, Sep 30 – Oct 3, 2004
26. IPAM Functional Genomics program reunion, Los Angeles, June, 2004
27. Department Seminar, Statistics, University of Chicago, Jan 2003
28. Department Seminar, Statistics, University of California at Berkeley, Feb 2003
29. Department Seminar, Statistics, Yale University, Feb 2003
30. Department Seminar, Statistics, University of California at San Diego, Feb 2003
31. Department Seminar, Biostatistics & Statistics, University of North Carolina at Chapel Hill, Mar 2003
32. Seminar, Genomic Center, North Carolina State University, Mar 2003
33. Department Seminar, Statistics, University of Connecticut, October 2002
34. Functional Genomics Conference, Los Angeles, June 2002
35. Celera Genomics, October 2001
36. Conference on Mathematical and Statistical Challenges in computational Biology
Lake Arrowhead, California, December 2000