Department of Statistics, 317 Evans Hall   Phone (510) 642-6433
University of California, Berkeley          E-mail hhuang@stat.berkeley.edu
Berkeley, CA 94720                          http://www.stat.berkeley.edu/~hhuang

# Haiyan Huang

| | |
|---|---|
| **Current Position** | **Associate Professor** with tenure (07/2010 – present) <br> **Assistant Professor** (07/2003-06/2010**)** <br><br> Department of Statistics <br> Interdepartmental Group in Biostatistics <br> Graduate Group in Computational and Genomic Biology <br><br> University of California, Berkeley, USA |
| **Education/ Training** | 9/2001-7/2003          Harvard University <br> **Post-Doctoral Fellow, Departments of Stat and Biostat (computational biology)** <br> Supervisors: Wing H. Wong, Jun S. Liu <br><br> 9/2000-12/2000          Institute for Pure and Applied Mathematics, UCLA <br> **Research Fellow in Functional Genomics** <br><br> 8/1997-8/2001          University of Southern California <br> **PhD, Applied Mathematics** <br> Advisor: Larry Goldstein <br> Dissertation: Bounds for the errors in word count distributional approximations. <br><br> 8/1993-7/1997          Peking University, China <br> **BS, Mathematics** |
| **Awards & Grants Received** | <ul><li>**Chau Hoi Shuen Foundation Women in Science Program (PI),** 7/2017-6/2018. "Computational approaches for diagnosing major psychiatric diseases."</li><li>**DOD CDMRP (BCRP Breakthrough Award) (co-I)** (PI: Herr), 2016-2019. "Scrutinizing ER Isoform Heterogeneity and Adverse Patient Outcomes in Triple Negative Breast Cancer." <u>My annual effort</u>: 0.5 summer month</li><li>**NIH/NCI 5R01CA203018-02 (co-I)** (PI: Herr), 12/2015 – 11/2019. "Multiplexed Isoform Quantification in HER2-Positive Breast Cancer." <u>My   annual effort</u>: 0.5 summer month</li><li>**Tsinghua-Berkeley Fund Award (PI)**, 7/2014-6/2016. "Modeling cis- and trans-interactions in gene expression and association networks"</li><li>**LBNL (NIH Prime; P01 7043194) (co-I)** (PI: Biggin), 10/2012-6/2017 "Quantitative Modeling of Transcriptional Information in the Drosophila Genome." <u>My annual effort</u>: 0.3-0.5 summer month</li><li>**NIH U01 HG007031 (co-PI)** (PI: Bickel), 9/2012-5/2017.  "Removing Statistical Bottle-Necks in Data Analysis for the Encode Consortium." <u>My annual effort</u>: 1 summer month + 1 graduate student (50%) + 1 postdoc (50%)</li><li>**NSF DMS-1160319 (co-PI)** (PI: Bickel), 6/2012-5/2016. "FRG: Collaborative</li></ul> |

Research: Unified Statistical Theory for the Analysis and Discovery of Complex Network." <u>My annual effort:</u> 1 summer month + 1 graduate student (50%) + 1 postdoc (50%)

- **NIH R21HG006187 (co-PI)** (PI: Bickel), 6/2011-5/2013. "Beyond Heuristics: A Tool for the Rigorous Statistical Analysis of *-Seq Assays."
  <u>My annual effort:</u> 1 summer month
- **NSF DMS-1026441 (co-PI)** (PI: Nolan), 2010-2013**.** "SCREMS: Building a Statistical Computing Environment to Support Scientific Research." Equipment funding
- **NIH R21EY019094 (PI)**, 6/2009-5/2012. "Applied Statistics to a Secondary Analysis of Public Repositories for Microarray Data"
- **EBI–UK (NIH Prime; ENCODE DAC) (co-PI)** (PI: Bickel), 2008-2013. "Develop Statistical and Computational Methods for Analyzing ENCODE Data." <u>My annual effort:</u> 1 summer month + 1 graduate student (50%) + 1 postdoc (50%)
- **NSF BECR11652 (co-PI)** (PI: Sohn), 2007-2010. "Collaborative Research: Nano-Cytometry: A Point-of-Care Technology for Monitoring Chronic Leukemia Patients." <u>My annual effort:</u> 0.5 summer month + 1 graduate student (50%)
- **USC (NIH Prime; 1R01GM074163) (subcontract PI)**, 1/2006-12/2010 "Integrative Analysis of Cross-Platform Microarray Data"
- **NIH 5R01GM075312 (co-PI)** (PI: Bickel), 6/2005-5/2009, "Comparative Genomics to Identify Functional Blocks and Horizontal Gene Transfer."
  <u>My annual effort:</u> 1 summer month + 1 graduate student (50%) + 1 postdoc (50%)
- **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

| | |
|---|---|
| **Research Interests** | Developing novel high-dimensional statistical methods and studying the underlying statistical theories to address modeling and analysis challenges from the cutting edge of biological and medical research involving big data. Topics I have examined include: |

- Statistical methodology developments and theoretical studies in areas including clustering, hierarchical multi-label classification, network analysis, measures of association in multivariate data, machine learning, distributional approximations, statistical optimization.
- Statistical genomics: *(i)* Identification of gene functions, relationships, pathways and networks using diverse, heterogeneous data, e.g., gene expression data (microarray or RNA-seq), regulatory data, snp data, etc. *(ii)* Single (cancer) cell data analysis; *(iii)* Prediction of functional elements in genome sequences.
- Translational bioinformatics: integrative information retrieval from diverse sources for (*i*) aiding disease diagnosis, *(ii)* predicting disease-drug relationship, *(iii)* revealing drug side effects, etc.
- Problems arising in other scientific fields with interesting data, e.g., protein profiling of circulating tumour cells (CTCs) by a microfluidic western blot, cell characterization by a Nanodevice, etc.

| **Publications** | (Students and postdocs supervised/co-supervised by me are highlighted in *blue italic* ) |
| --- | --- |
| | (**\*** denotes senior authors ) |
| | (**\$\$** denotes that I serve as the only senior statistician on a collaborative project) |

1. *Wang YXR*, *Liu K*, Theusch E, Rotter JI, Medina MW, Waterman M**\***, **Huang H\*** (2017) Generalized correlation measure using count statistics for gene expression data with ordered samples. Revision submitted to **Bioinformatics.**

2. *Pimentel H, Hu ZT*, **Huang H\*** (2017). Biclustering by Sparse Canonical Correlation Analysis. **Quantitative Biology**. Accepted.

3. **Huang H**, Yu B (2017). Data Wisdom in Computational Genomics Research. **Statistics in Biosciences.** 2017: 1-16.

4. Chang M, Womer FY, Edmiston KE, Bai C, Zhou Q, Jiang X, Wei S, Wei Y, *Ye Y*, **Huang H\$\$**, He Y, Xu K, Tang Y, Wang F (2017). Neurobiological Commonalities and Distinctions Among Three Major Psychiatric Diagnostic Categories: A Structural MRI Study. **Schizophrenia Bulletin.** 2017 Jun 13.

5. *Shi F*, **Huang H\*** (2017). Identifying Cell Subpopulations and Their Genetic Drivers from Single-Cell RNA-Seq Data Using a Biclustering Approach. **Journal of Computational Biology.** 2017 Jul 1;24(7):663-74.

6. *Schiffman C*, Lin C, *Shi F*, Chen L, Sohn L, **Huang H\*** (2017). SIDEseq: A Cell Similarity Measure Defined by **S**hared **I**dentified **D**ifferentially **E**xpressed Genes for Single-cell RNA Sequencing data. **Statistics in Biosciences.** 2017: 1-17.

7. Sinkala E, Sollier-Christen E, Renier C, Rosàs-Canyelles E, Che J, Heirich K, Duncombe TA, Vlassakis J, Yamauchi KA, **Huang H\$\$**, Jeffrey SS, Herr AE (2017). Profiling protein expression in circulating tumour cells using microfluidic western blotting. **Nature Communications**, Mar 23;8:14622.

8. *Wang YXR*, Jiang K, Feldman L, Bickel PJ, **Huang H\*** (2015). Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. **Annals of Applied Statistics.** 9(1): 300-323.

9. National Research Council Committee on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures (2015). "**National Academy of Sciences: Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense**." Washington, DC: The National Academies Press.
   (I served as an NRC Committee member)

10. *Wang YXR*, Waterman M**\***, **Huang H\*** (2014). Gene coexpression measures in large heterogenous samples using count statistics. **Proc Natl Acad Sci USA.** 111(46):16371-6

11. *Kim K*, Bolotin E, Theusch E, **Huang H\$\$**, Medina MW, Krauss RM (2014). Prediction of LDL cholesterol response to statin using transcriptomic and genetic variation. **Genome Biology.** 15(9): 460.

12. *Jiang CR*, Liu CC, Zhou XJ, **Huang H\*** (2014)**.** Optimal Ranking in Multi-label Classification Using Local Precision Rates. **Statistica Sinica**. 24: 1547-1570.

13. Gerstein M, et al. (2014). Comparative Analysis of the Transcriptome across Distant Species. **Nature.** 512: 445-448. doi:10.1038/nature13424.

14. Snyder M, et al. (2014). Comparative analysis of regulatory information and circuits

across distant species. **Nature.** 512: 453-456. doi:10.1038/nature13668.

15. *Li JJ*, **Huang H\*,** Bickel PJ\*, Brenner SE\* (2014). Comparison of D. melanogaster and C elegans developmental stages, tissues, and cells by modENCODE RNA-seq data. **Genome Research.** 24: 1084-1101.

16. *Wang YXR*, **Huang H\*.** Review on statistical methods for gene network reconstruction using expression data (2014). **Journal of Theoretical Biology.** 362:53-61.

17. *Xiong H*, Brown BJ, Boley N, Bickel P, **Huang H\*** (2014). DE-FPCA: Testing Gene Differential Expression and Exon Usage through Functional Principal Component Analysis. In S. Datta and D. Nettleton, editors, "**Statistical Analysis of Next Generation Sequence Data (Frontiers in Probability and Statistical Science).**" Springer. New York. (ISBN-13: 978-3319072111 | ISBN-10: 3319072110)

18. *Li JJ*, **Huang H\*,** Qian M, Zhang X (2014). Transcriptome analysis using next-generation sequencing. In Y. Lu, L. Tian, J. Fang, H. Jin, editors "**Advanced Medical Statistics**", 2nd Edition. (Publication Date: Dec 20, 2014 | ISBN-10: 9814583294 | ISBN-13: 978-9814583299)

19. Chapman MR, Balakrishnan KR, Li J, Conboy MJ, **Huang H$$**, Mohanty SK, Jabart E, Hack J, Conboy IM, Sohn LL (2013). Sorting single satellite cells from individual myofibers reveals heterogeneity in cell-surface markers and myogenic capacity. **Integrative Biology**. 5(4):692-702.

20. ENCODE Consortium Project (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. **Nature.** Nature. 2012 Sep 6;489(7414):57-74.

21. *Kim K*, Jiang K, Teng S, Feldman LJ, and **Huang H\*** (2012). Using biologically interrelated experiments to identify pathway genes in Arabidopsis. **Bioinformatics**, 28(6), 815-822.

22. *Gao Q, Ho C*, Jia Y, *Li JJ*, **Huang H\*** (2012). Biclustering of Linear Patterns in Gene Expression Data (CLiP). **Journal of Computational Biology**, 19(6), 619-631.

23. Chapman MR, Balakrishnan K, Conboy MJ, Mohanty SK, Jabart E, Li J, **Huang H$$**, Hack J, Conboy IM, and Sohn LL (2012). Label-free screening of niche-to-niche variation in Satellite stem cells using functionalized pores. A chapter in **Nanopores for Bioanalytical Applications: Proceedings of the International Conference**, Eds. J. Edel and T. Albrecht, RSC Publishing, pp. 38-42.

24. *Li JJ, Jiang CR,* Brown JB, **Huang H\*,** Bickel PJ\* (2011). Sparse Linear Modeling of RNA-seq Data for Isoform Discovery and Abundance Estimation. **Proc Natl Acad Sci. USA**, 108 (50) 19867-19872.

25. *Li Q*, Brown JB, **Huang H** and Bickel PJ (2011). Measuring reproducibility of high-throughput experiments. **Annals of Applied Statistics**. 5(3), 1752-1779.

26. Li Y, **Huang H** and Cai L (2011) Prediction of Transcriptional Regulatory Networks for Retinal Development. A chapter in book **Computational Biology and Applied Bioinformatics.**

27. Durinck S, *Ho C*, Wang NJ, Liao W, Jakkula LR, Collisson EA, Pons J, Chan SW, Lam ET, Chu C, Park K, Hong S, Hur JS, Huh N, Neuhaus IM, Yu SS, Grekin RC, Mauro TM, Cleaver JE, Kwok P, LeBoit PE, Getz G, Cibulskis K, Aster JC, **Huang H**, Purdom E, Li J, Bolund L, Arron ST, Gray JW, Spellman PT, Cho RJ (2011). Temporal Dissection of Tumorigenesis in Primary Cancers. **Cancer Discovery**, 1:137-143.

28. *Xu N*, Bickel PJ, **Huang H\*** (2010). Genome-wide Detection of Transcribed Regions

through Multiple RNA Tiling Array Analysis. **International Journal of Systems and Synthetic Biology**, 1(2) 155-170.

29. Bickel PJ, Boley N, Brown JB, **Huang H**, Zhang NR (2010). Subsampling Methods for Genomic Inference. **Annals of Applied Statistics.** Vol. 4, No. 4, 1660-1697.
(The authors were alphabetically ordered; equal contribution)

30. **Huang H\*,** Liu C, Zhou XJ\* (2010). Bayesian Approach to Transforming Public Gene Expression Repositories into Disease Diagnosis Databases. **Proc Natl Acad Sci. USA**. 107 (15) 6823-6828.
    - *Under the title "Gene databases mined for diagnoses", this paper was selected as an issue highlight by PNAS: http://www.pnas.org/content/107/15/6553.full.pdf+html*
    - *It was selected by the Faculty of 1000 Biology (http://www.f1000biology.com) and evaluated by Russ Altman (Stanford University): http://www.f1000biology.com/article/id/3925957.*
    *(Faculty of 1000 Biology is an award-winning online service that highlights and evaluates interesting biology papers, based on the recommendations of over 2000 of the world's top researchers.)*
    - *It was featured in the following news reports:*
    *(i) GenomeWeb daily news. "Team Develops Proof-of-Principle Diagnostic Database for Applying Public Gene Expression Data". March 22, 2010.*
    *(ii) National Cancer Institute Research News. "Mathematical Modeling Turns Gene Expression Data into Disease Diagnostics" http://physics.cancer.gov/news/2010/april/po_news_b.asp*
    *(iii) Biocentury and Nature publishing group. "GEO: world of diagnostic potential," Haas, M.J. SciBX 3(14); doi:10.1038/scibx.2010.426. April 8, 2010*

31. Jiang K, Zhu T, Diao Z, **Huang H$$**, Feldman LJ. (2010). The Maize Root Stem Cell Niche: A Partnership between Two Sister Cell Populations. **Planta,** 231(2):411-24.

32. 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** 367, 4313-4337.
(Authors were ordered alphabetically)

33. *Teng SL*, **Huang H\*** (2009). A Statistical Framework to Infer Functional Gene Associations from Multiple Biologically Interrelated Microarray Experiments. **Journal of the American Statistical Association**, June 2009, Vol. 104, No. 486.

34. 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 anterio cingulum integrity in patients with schizophrenia and healthy controls. **Journal of Psychiatry & Neuroscience,** 2009 May;34(3):181-6.

35. 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.

36. 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.

37. **Huang H**, Cai L, Wong WH. (2008) Clustering analysis of SAGE transcription profiles using a Poisson approach. **Methods in Molecular Biology**. 2008; 387:185-98.

38. 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.

39. *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*).

40. ENCODE Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature.** 447, 799-816.

41. **Huang H**, *Kim K* (2006). Unsupervised clustering analysis of gene expression. **Chance,** vol. 19, No.3.

42. 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.

43. 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.

44. 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)

45. **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.

46. 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)

47. Zhao X, **Huang H,** Speed T (2004). Finding short DNA motifs using permuted Markov models. **Proceedings of RECOMB 2004.**

48. 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 Biology**, 2(9):E247.

49. Allinen M, Beroukhim 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.

50. Lippert RA, **Huang H**, Waterman MS (2002). Distributional regimes for the number of *k*-word matches between two random sequences. **Proc Natl Acad Sci. USA,** 99(22):13980-9.

51. **Huang H** (2002). Error bounds on multivariate normal approximations for word count statistics. **Advances in Applied Probability,** 34(3): 559-586.

**Submitted Manuscripts**

52. Chang M, Edmiston KE, Womer FY, Zhou Q, Wei S, Jiang X, *Ye Y*, **Huang H[$$]**, Xu K, Tang Y, Wang F (2017). Spontaneous Low-Frequency Fluctuations in Neural System for Emotional Perception in Major Psychiatric Diagnostic Categories: Amplitude Similarities and Differences across Frequency Bands. Submitted.

53. Kang CC, Ward TM, Bockhorn J, *Schiffman C*, **Huang H[$$]**, Pegram MD and

Herr AE (2017). Clinical Electrophoretic Pathology Investigates Oncoprotein Isoforms. Submitted to NPJ (Nature Partnering Journal) Precision Oncology.

| | |
|---|---|
| **Manuscripts in Preparation** | 54. *Liu K,* Theusch E, Rotter JI, Bickel PJ\*, Medina MW\*, **Huang H\*** (2017). Bagged semi-supervised gene clustering for context-specific gene pathway analysis. |

55. *Ho C, Jiang CR, Lee W,* **Huang H\*** (2017). Optimal decision making in hierarchical multi-label classification.

56. Ren H, Qi C, Ma Q, Sun X, Wei Q, *Liang JW,* Li G, Zhang Z, Ren J, Yan J, Li G, Xu W, Li X, Bai C, She X, Wu G, **Huang H$^{\$\$}$** and McCormick S (2017). Fixation of expression divergences by natural selection in *Arabidopsis* coding genes.

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

• **"Inaugural DahShu Data Science Symposium: Computational Precision Health (CPH 2017)",** edited by Haiyan Huang, Hua Tang, et. al.. Special Issue published by the **Journal of Computational Biology**.

| | |
|---|---|
| **Software Developed** | (The paper indices in superscript refer to those in the publication list) |

***GEA[39,41,46]:*** For clustering/classification/significance analysis of SAGE and Microarray gene expression data (http://cell.rutgers.edu/gea)

(*First model-based method for clustering analysis of SAGE data*)

***Knorm[33]:*** For gene association inference across multiple dependent experimental conditions. (http://cran.r-project.org/web/packages/knorm)

(*It is for concurrently estimating gene-gene dependencies and experiment-experiment dependencies in a gene expression matrix; we have extensive evidence to show the importance and necessity of considering such dual dependencies in analyzing gene expression data*)

***LMM[45]:*** For predicting transcription factor binding sites by evaluating a candidate site in a local genomic context. (http://www.stat.berkeley.edu/users/hhuang/LMM.zip)

***TilingAnalyzer[28]:*** For transcriptome detection by analyzing multiple RNA tiling arrays (http://www.stat.berkeley.edu/users/hhuang/TilingAnalyzer.zip)

***CLiP[22]:*** For biclustering analysis of gene expression data; detecting genes that are linearly correlated in a subset of samples (http://www.stat.berkeley.edu/users/hhuang/CLiP.zip)

***SCCAB[2,8]:*** For biclustering using sparse canonical correlation analysis (github.com/pimental/scaab)

***SIDEseq[6]:*** For measuring cell similarity; facilitating the identification of subpopulations of cells based on single cell RNA-seq data. R codes are available upon request.

**Professional Activities**

- **Associate Editor** for the Journal of American Statistical Association (JASA), Theory & Methods, 2017 - present
- **Associate Editor** for the Annals of Applied Statistics, 2015 - present
- **Editorial Board Member**, Journal of Computational Biology, 2014-present
- **VP of Scientific Outreach** & co-founder, Dahshu, non-profit organization to promote research and education in data sciences, 2016 - present
- **Organization Committee co-Chair,** DahShu 2017 conference: Data Science and Computational Precision Health, San Francisco, Feb 20-22, 2017
- **Serving** on the NSF/MCB review panel, 2017
- **Guest Editor** for PLOS Computational Biology, October, 2016
- **Member of the Committee on Nominations of the Institute of Mathematical Statistics (IMS),** 2015-2016; the duty was to find a new president-elect for IMS, and to nominate new Council members
- **Local Committee co-Chair,** APBC 2016, Jan 10-13, San Francisco, California
- **Serving** on the NSF/NIH BIGDATA review panels in 2012-2015
- **Member of the National Research Council (NRC) Committee** on Predictive-Toxicology Approaches for Military Assessments of Acute Exposures in 2014-15
- **Participation in the Statistics advisory work group** to clarify the course criteria and faculty expectations for California Community College (CCC) courses, organized by Board of Admissions & Relations with Schools (BOARS)
- **Organizer of the invited session** on "Statistical Networks in Biology: Beyond the Standard Paradigm," Australian Statistical Conference in conjunction with the Institute of Mathematical Statistics Annual Meeting, Sydney, July 2014
- **Serving** on the NSF DMS statistics review panel in 2013
- **Program Committee Member**, Annual International Conference on Intelligent Systems for Molecular Biology (ISMB), 2010-2016
- **Associate Editor** for "Statistical Applications in Genetics and Molecular Biology", 2004-2013.
- **Organizer of the invited session** on "Translational Bioinformatics: Secondary analysis of high-dimensional public genomic data", Annual Joint Statistical Meeting, 2012
- **Serving** on the NIH study Section "The Genomics, Computational Biology and Technology" in 2011
- **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 national conference in Statistics; I organized 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. (An international conference in Probability and Statistics)
- **Participant** in the NSF/CAUSE-sponsored workshop on "Integrating Computing i 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.
- **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, 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.

| Additional Professional Experience | **Department and Campus-wide Academic Services** |
|---|---|
| | **Teaching Committee Chair,** Department of Statistics, 2013-2014, 2015-present |
| | **Chair of PhD Program,** Center of Computational Biology, UC Berkeley, 2017-present |
| | **Head Graduate Advisor,** Center of Computational Biology, UC Berkeley, 2016-present |
| | **Participate as a Berkeley Science Network faculty leader** at the BSN-BSC Deep Dive Retreat, Feb 2014, April 2017 |
| | **Member of the benchmarking committee** that established acceptable scores on the Versant English Test for GSI appointments, UC Berkeley, May 2015 |
| | **Reviewer for France-Berkeley Fund,** 2012, 2014 |
| | **Participant of the How Students Learn: Spring Seminar and Summer Funding Program**, sponsored by the Graduate Council's Advisory Committee for GSI Affairs, 2014 |
| | **Faculty Hiring Committee,** Department of Statistics, 2012-2013 |
| | **Consulting and Outside Matters Committee,** Department of Statistics, Fall 2010 |
| | **Statistics Certificate/Professional MA Exploration Committee member,** Department of Statistics, 2010-2011 |
| | **Organizer for Faculty Research Lectures**, Department of Statistics, Fall 2010 |

**Course Committee member,** Department of Statistics, Fall 2009, 2010, 2011
**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, 2008-9
**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, Spring 2010
**Student Admission Committee Member**, Department of Statistics, 2004-2006
**Organizer** of Statistics women faculty/students lunch meeting, Spring 2004
**Colloquium Organizer**, Stat Dept, UC Berkeley, 2003-2005, Spring 2008, 2009

## Teaching

Undergraduate courses
STAT 152: Survey Sampling (*Fall 2003, Fall 2004, Fall 2005, Fall 2006*)
BIOE/STAT C141: Statistics for Bioinformatics (*Springs 2004, 2005, 2006, 2007, 2008*)
STAT 131A: Statistical Inferences for Social and Life Scientists (*Spring 2009*)
STAT 157: Seminar on Topics in Probability and Statistics (*Fall 2009*)
STAT 133: Concepts in Computing with Data (*Spring 2013*)

Master courses
STAT 200B: Introduction to Probability and Statistics at an Advanced Level (*Springs 2006, 2007, 2011, 2012*)
STAT 201B: Introduction to Probability and Statistics at an Advanced Level (*Falls 2013, 2014, 2016*)

PhD courses
STAT 215B: Statistical Models: Theory and Application (*Spring 2017*)
STAT 215A: Statistical Models: Theory and Application (*Fall 2011*)
STAT 210A: Theoretical Statistics (*Falls 2008, 2009, 2010*)
STAT 246: Statistical Genetics (*Spring 2009*)
STAT C245E,F: Statistical Genomics (*Springs 2010, 2012, 2013, 2014*)
STAT 272: Statistical Consulting (*Fall 2010, Spring 2014, Fall 2014, Spring 2015, Spring 2016*)

Pedagogical course
STAT 375: Professional Preparation: Teaching of Probability and Statistics (*Falls*

*2013, 2014*)

**Advising**

### Current Graduate Students

Courtney Schiffman (Biostat PhD program, UC Berkeley, 09/2015 – present)
Yuting Ye (Biostat Master, Biostati PhD Program, UC Berkeley, 2016 – present)
Yun Zhou (Statistics Master, Biostatistics PhD program, UC Berkeley, 2016 – present)
Daisy Yan Huang (Statistics PhD program, UC Berkeley, 03/2008 – present)
Funan Shi (Statistics PhD program, UC Berkeley, 09/2012 – present)

### Current Postdoc

Ke Liu (joint with Prof. Peter Bickel, 03/2015 – present; PhD in 2015, Tsinghua University)

### Former Postdocs

Hao Xiong (Postdoc, joint with Prof. Peter Bickel, 01– 12/2012; PhD in 2009, Computer Sci., Texas A&M U)
Current position: Research fellow, UW Seattle
Qunhua Li (Postdoc, joint with Prof. Peter Bickel, 09/2008 – 07/2011; PhD in 2008, Statistics, UW Seattle)
Current position: Assistant Professor, Statistics, Penn State U

Ci-Ren Jiang (Postdoc, 09/2009 – 09/2010; PhD in 2009, Statistics, UC Davis)
Current position: Assistant Research Fellow, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

### Former PhD Students

Christine Ho (PhD in 2016). Immediate Position After Graduation: Statistical Analyst in Pandora.
Rachel Wang (PhD in 2015; co-advised with Prof. Peter Bickel). Immediate position after graduation: Stein Fellow of Statistics, Stanford University. Will join the faculty of Statistics in the University of Sydney, Australia from July 2017.
Jessica Li (PhD in 2013; co-advised with Prof. Peter Bickel). Immediate position after graduation: Assistant Professor, Statistics, UCLA.
Kyungpil Kim (PhD in 2013). Immediate position after graduation: Postdoc at Children's Hospital Oakland Research Institute.
Na Xu (PhD in 2008; co-advised with Prof. Peter Bickel). Immediate position after graduation: statistician at Genentech Inc.
Siew-Leng Melinda Teng (PhD in 2007). Immediate position after graduation: statistician at Genentech Inc.

### Former Master Students

Harold Pimentel (Master in Dec 2013). Immediate position after graduation: PhD student in Computer Science, UC Berkeley.

Meng Ling (Master in 2010). Immediate position after graduation: research fellow at UC Berkeley, Department of Plant and Microbial Molecular Biology

Hua Chen (Master in 2008). Immediate position after graduation: research fellow at Harvard Univ, Department of Genetics

**Former Undergraduate Students**

Phillip Rescober (Undergraduate 2004; honor thesis advising)

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

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

Dan Nguyen (Undergraduate 2007), RA, summer 2007

Qui Tran (Undergraduate 2010; honor thesis advising)

**Former Visitors**

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

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

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

**Major Collaborators**

Peter Bickel research group (Statistics, UC Berkeley)

Amy Herr (Bioengineering, UC Berkeley)

Lydia Sohn research group (Mechanical Engineering, UC Berkeley)

Ron Krauss and Marisa Medina research group (Children's Hospital at Oakland Research Institute)

Lin He research group (Molecular and Cell Biology, UC Berkeley)

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

**Invited talks**

1. (To be given) AIM workshop on "Stein's method and applications in high-dimensional statistics", August 6-10, 2018
2. (To be given) Featured presentation, 2018 Center for Genome Research and Biocomputing (CGRB) Spring Conference/Colloquium at Oregon State University, April, 2018
3. (To be given) ENAR annual statistical meeting, Atlanta, March 25-28, 2018
4. (To be given) IMS workshop on "Meeting the Statistical Challenges in High Dimensional Data and Complex Networks", Singapore, Feb 5-16, 2018
5. (To be given) Department seminar, Statistics, Carnie Mellon U, December 6, 2017
6. (To be given) Department seminar, Statistics, U Chicago, December 4, 2017
7. Joint Statistical Meeting, Baltimore, August 3rd, 2017
8. Seminar, Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China, May 2017
9. Department seminar, Bioinformatics Division, TNLIST, Tsinghua University, Beijing, China, May 2017
10. DahShu 2017 conference: Data Science and Computational Precision Health, San

Francisco, Feb 20-22, 2017

11. Department seminar, Bioinformatics Division, TNLIST, Tsinghua University, Beijing, China, August 2016
12. Seminar, Cancer Research UK Cambridge Institute, England, July 2016
13. Isaac Newton workshop: Graph Limits and Statistics, England, July 2016
14. Department seminar, Biostatistics, UC San Francisco, May, 2016
15. Short course on *Modern Statistical Methods for Big Data and Big Questions in Genomics* (joint with Dr. Sandrine Dudoit and Dr. Elizabeth Purdom), APBC 2016 San Francisco, January 2016
16. Berkeley-Stanford Statistics joint colloquium, Stanford University, Nov 2015
17. Biology and Mathematics in the Bay Area (BaMBA) annual meeting, Oct 2015
18. Joint Statistical Meeting, Seattle, August 2015
19. ENAR Annual Statistical Meeting, Miami, March 2015
20. Specially-Invited Speaker in the sixth National Bioinformatics meeting, Nanjing, China, October 2014
21. Joint Statistical Meeting, Boston, August 2014
22. ASC-IMS annual meeting, Sydney, Australia, July 2014
23. Tsinghua Chair Professors' Bioinformatics Workshop, Beijing, China, May 2014
24. International Workshop on Computational Biology and Bioinformatics, Fudan University, Shanghai, China, May 2014
25. Department Seminar, Statistics, UC Riverside, November, 2013
26. Department Seminar, Biostatistics, UC Davis, October, 2013
27. Keynote speaker in the IEEE International Conference on Systems Biology (ISB 2013), Huang Shan, China, Aug 2013
28. Department seminar, Shanghai Institutes for Biological Science, China, Aug 2013
29. Department seminar, China Medical University, Shenyang, China, Aug 2013
30. WNAR statistics meeting, UCLA, June 2013
31. Joint Statistical Meeting, San Diego, July 2012
32. Department Seminar, Statistics, Peking University, July 2012
33. Invited Seminar, MOE Key Laboratory of Bioinformatics, Tsinghua University, China, July 2012
34. Computational Biology Symposium, USC, Mar 2012
35. Department Seminar, Statistics, University of Chicago, Mar 2012
36. Invited Seminar, Children's Hospital Oakland Research Institute (CHORI), Jan, 2012
37. Department Seminar, Biostatistics, Stanford University, Nov 2011
38. Joint Statistical Meeting (contributed session), Miami, July 2011
39. Department Seminar, Statistics, Yale University, Feb 2011
40. Guest Lecturer, Statistics 166/366 (Statistical Models in Biology), Department of Statistics, Stanford University, May 2010
41. Department Seminar, Statistics, University of Wisconsin-Madison, March 2010
42. Department Seminar, Statistics, Stanford University, October 2009
43. Workshop on "Mathematical Genomics", Mathematical Sciences Research Institute, April 13 – 15, 2009.
44. Department Seminar, Biostatistics, UCSF, Oct 2008

45. WNAR/IMS Joint Annual Meeting, Davis, June 2008
46. Department Seminar, Statistics, UC Davis, June 2008
47. Newton Institute Workshop on "High Dimensional Statistics in Biology", Cambridge, UK, March 31- April 4, 2008
48. Department Seminar, Statistics, University of Illinois at Urban- Champaign, Oct 2007
49. **Keynote speaker** for the workshop/conference on "Statistics for Biomolecular Data Integration and Modeling," Monte Verita, Ascona, Switzerland, June 2007
50. Department Seminar, Center of Statistical Sciences, Brown University, April 2007
51. Research Seminar, GENETECH, January 2007
52. Biostatistics Seminar, UCSF, November 2006
53. Department Seminar, Statistics, Stanford University, October 2006
54. ICSA 2006 Applied Statistics Symposium, University of Connecticut, June 14-17, 2006
55. Workshop on Frontiers of Statistics, Princeton University, May 18-20, 2006
56. Workshop on "Sequence Analysis Toward System Biology", IPAM, Jan 2006
57. Statistics Department Seminar, UCLA, Oct 2005
58. Department Seminar, Computational Molecular Biology and Bioinformatics, USC Oct 2005
59. Statistics Department Seminar, Duke University, Sept 2005
60. Joint Statistical Meeting, Minnesota, Aug 2005
61. IMS/CSPS Joint Meeting, Beijing, China, July 2005
62. WNAR/IMS Joint Meeting, Alaska, June 2005
63. Biostatistics Seminar, SCHARP/FHCRC, Seattle, May 2005
64. Department Seminar, Statistics, University of California at Davis, May 2005
65. Department Seminar, Biostatistics, Stanford University, November 2004
66. Department of Biostatistics, University of Pittsburgh. October 2004
67. SAGE 2004 conference on decoding the "OMEs", Boston, Sep 30 – Oct 3, 2004
68. IPAM Functional Genomics program reunion, Los Angeles, June, 2004
69. Department Seminar, Statistics, University of Chicago, Jan 2003
70. Department Seminar, Statistics, University of California at Berkeley, Feb 2003
71. Department Seminar, Statistics, Yale University, Feb 2003
72. Department Seminar, Statistics, University of California at San Diego, Feb 2003
73. Department Seminar, Biostatistics & Statistics, University of North Carolina at Chapel Hill, Mar 2003
74. Seminar, Genomic Center, North Carolina State University, Mar 2003
75. Department Seminar, Statistics, University of Connecticut, October 2002
76. Functional Genomics Conference, Los Angeles, June 2002
77. Celera Genomics, October 2001
78. Conference on Mathematical and Statistical Challenges in computational Biology Lake Arrowhead, California, December 2001