SUSMITA DATTA (CV Revised 05/24/2019)

Advisory board member of the NSF/CBMS Regional Research Conferences in Mathematics 2018-

Featured women statistician in March 2019 issue of Amstat News in the honor of Women's History Month "Celebrating Women in Statistics and Data Science"

Featured in Amstat News in June 2019 issue "oLet's Chat: Top WSDS Speakers Provide Insights Into Their Careers"
This year's WSDS keynote speaker is Donna Brogan, and the plenaries will be from Jeri Mulrow, Bonnie Ray, and Susmita Datta. Here, they reflect on their long careers, share advice for the future, and discuss the topics they plan to talk about.

Featured CWIS President in Statistics Views at Wiley.com 23 Oct 2014

Fellow of American Association for the Advancement of Science (AAAS) (2014)

President of Caucus for Women in Statistics (2013)

Fellow of American Statistical Association (ASA) (2012)

Elected Member of International Statistical Institute (ISI) (2010)

H-index 31

Preeminent Hire, Professor (tenured), Department of Biostatistics, University of Florida, Gainesville, FL (2015 Fall -
Department of Biostatistics
2004 Mowry Rd., Rm 5226 CTRB
Gainesville, FL 32610, USA
(352)294-5923
susmita.datta at ufl dot edu

Co-Director of The Biostatistics, Epidemiology and Research Design Program (BERD) of Clinical Translational Research Institute (CTSI) of the University of Florida (2016-)

Adjunct Professor and Distinguished University Scholar (2015 Fall-
Department of Bioinformatics & Biostatistics School of Public Health and Information Sciences University of Louisville
485 East Gray Street
Louisville, KY 40292

EDUCATION
1995	PhD	University of Georgia (Statistics)
1990	MS	University of Georgia (Applied Statistics)
1986	BS	University of Calcutta, India (Physics)

PREVIOUS POSITIONS

2005 - 2015   @ University of Louisville, Louisville, KY

2012 (August) – 2015 (May) Graduate Program Director, Department of Bioinformatics and

Biostatistics (MS and PhD)
2013 (June) – 2015 (August) Distinguished University Scholar
2010 (June) – 2013 (May) University Scholar
2010 (January) - Professor (Tenured), Department of Bioinformatics & Biostatistics
2005-2009   Associate Professor (Tenured), Department of Bioinformatics & Biostatistics

1997 - 2005 Georgia State University, Atlanta, GA

2002-2005   Associate Professor (Tenured), Department of Mathematics and Statistics
2002-2005   Associate Professor, Department of Biology (joint appointment)
1997-2002   Assistant Professor, Department of Mathematics and Statistics

1995-1997 Emory University, Atlanta, GA

1995-1997   NRSA Post- Doctoral Fellow, Department of Biostatistics

2000-2001 University of Georgia, Athens, GA

2000(Fall) - 2001(Summer) Visiting Assistant Professor, Department of Genetics.

RESEARCH INTERESTS

Methodological: Bioinformatics, Clustering and Classification, Genomics, Proteomics, Infectious Disease Modeling, Non-linear Regression modeling for Systems Biology, Statistical Issues in Population Biology, Statistical Genetics, Systems Biology, Survival Analysis and Multi state models.

Clinical: Birth Defects Research; Cancer Research

PHYLOSOPHICAL INTERESTS

Promoting women in STEM fields; Practicing Random Act of Kindness

RESEARCH GRANTS AND CONTRACTS

PI level

Pending:
1. R01HG010419-01 (Datta) Comprehensive Statistical Inference for Single Cell RNA seq Data 09/01/2018-08/31/2022 2.0 Calendar
    Direct Cost: $1,171,249.00

Current:
2. Project/Proposal Title: University of Florida Start up fund (PI)
    04/01/2016 – 3/31/2019
    Source of Support: University of Florida
    Total direct cost: $150,000
    Person-Months Per Year Committed to the Project. NA *UL1TR001427*

3. PI R15, 20% effort, National Institutes of Health, NCI 1R15CA170091-01A1. Identification of Proteins from Mass Spectrometry Data: A Statistical Approach 2013-2017 (last year NCE). Direct Cost $463,508.

4. PI R13,"Conference for Women in Statistics" National Institutes of Health, NIGMS R13 GM109727-01. 2014-2015. Direct Cost: $32,000

5. PI (no co-I) R15, 25% effort, National Institutes of Health, NCI R15 CA133844-01. Development of Statistical Methods for Analyzing Proteomic Cancer Data, 2009-2012. Direct Cost: $463,508.00

6. PI (no co-PI) NSF, 17% effort, National Science Foundation, Statistics Program (DMS), Standard Grant, Statistical peak detection, adaptive classification and protein-protein network construction using mass spectra, DMS-0805559, 2008-2011. $149,391.00

7. PI U of L subcontract (E. Voit, PI, Georgia Tech.) 10% effort, The Trehalose Cycle as Paradigm, National Science Foundation, 2006-2009.

8. PI 30% effort, Pilot Project, Proteomics Based Approach for Early Detection of Fetal Alcohol Syndrome, P20-RR/DE17702, NIH COBRE (PI, R. Green), 2006-2007. Direct cost: $50,000

9. PI (no co-PI) NSF (DMS) Statistical Analysis of Microarray Gene Expression Data, $127,671, 2000-2002.

10. PI (no co-PI), Research Experience for Undergraduates in Fungal Genomics and Computational Biology: GSU VPRSP grant, $18,666, Summer 2001.

11. PI (no co PI) A Pilot Project for Developing Statistical Tools for Bioinformatics. GSU faculty initiation grant, $5000, 2000-2001.

Non-PI level

Pending:

1. R01CA233604-01 (Yaghjyan) Air Pollution, Greenness, and Breast Cancer Risk in Breast Cancer Surveillance Consortium 9/1/2018 to 8/31/2021, 1.2., 1.8 Calendar
Direct Cost: $1,018,164.00

(Current)
1. 2P30AG028740-11 (Co-Investigator) 06/01/17 – 03/31/2022
Title: University of Florida Claude D. Pepper Older Americans Independence Center
Source of Support: NIH
Goal: Narrative Better knowledge of the factors leading to optimizing mobility and developing efficacious and effective interventions to avert disability will have a major impact on practicing evidence-based geriatric medicine and public health.
Person-Months Per Year Committed to the Project. Cal: 0.24
Total Direct Cost: $709,892

2. 1-596002052-A1 (Biostatistician) 09/01/2015-08/31/2020
Title: Membranes of the Dental Pathogen Streptococcus Mutans
Source of Support: NIH
Goal: Multiple virulence properties of *S. mutans* stem from the way the organism inserts proteins into its cytoplasmic membrane, but the cellular machinery that transports these proteins is not completely understood. This research will explain how proteins important for disease get into and through the membrane of *S. mutans* and other related pathogens.
Person-Months Per Year Committed to the Project. Cal: 0.6
Total Direct Cost: $ $1,875,000.00

3. Title: Clinical Metabolomics

Source of Support: NIH
Goal: Statistical Help to clinical metabolomics projects
Role: Key Personnel
Person-Months Per Year Committed to the Project. Cal: 0.53
Current year direct cost: $196,677.83

4. UL1TR001427 04/01/2016 – 3/31/2019
    Source: NIH
    Title: CTSI NIH Biostatistics Epidemiology Research Design (BERD) Program
    (Co-Director, Key Personnel)
    Goal: Provide statistical support and direct the research design studio
    Person-Months Per Year Committed to the Project. Cal: 1.06
    Current year direct cost: $156,409.42

5. Training Program for Applied Research and Development in Genomic Medicine (Johnson)
    Source: National Institute of Health (NIH)
    Dates: 09/01/2017-08/30/2022
    Role: Faculty mentor/Executive Member

    (Past)

1. Co-I R01 NIDDK/NIH 10% effort biomarker validation and dosing algorithms for anemia management in ESRD 12/01/2011-11/30/2016. Direct Cost: $590,663.00

2. Co-I R01, 8% effort, NIDDK/NIH DNA Sequences Impact Estrogen and Anti-Estrogen Activities, NIH R01 DK053220-10A1, (Carolyn M. Klinge, PI) 2008-2013.

3. Co-I R01, 8% effort, 01/01/10-12/31/14 NCI/NIH, Regulation of miRNA in breast cancer. 1R01CA138410-01A1 (PI Klinge, Carolyn M.), 2/1/2009-1/31/14.
    Direct Cost: $305,497

4. Co-I, R01 U of Louisville Subcontract, 5% effort, NIH, Environmental Interactions with the Genome and Epigenome. 2008-2010.

5. (Role: Co-I, Mentor), NCRI 3P20RR016481- 09S1 KY IDEA Networks of Biomedical Research Excellence, RC2 proposal. (Cooper, PI), 4/1/2009-3/31/2011

6. Co-director of Biostatistics Consortium for the Research Development Core 15% effort submitted CTSI proposal, 2010.

7. Biostatistics Advisor on Bioinformatics core, 10% effort submitted CTSI proposal, 2010.

8. Co-I/Biostatistician, 10% effort, NIH P01 AA017103-01, Alcohol, Liver Disease and Alcohol Nutrient Interactions, (Craig McClain, PI), 2008-2013.

9. Biostatistics Group Leader, Bioinformatics, Biostatistics and Computational Biology Core, Center for Environmental Genomics and Integrative Biology, P30ES014443-01A1 (K. Ramos, PI), 10% effort, NIEHS-NIH, 2007-2011.

10. Co-I (M. J. Kennedy, PI, Louisville) 5% effort, Aminoglycoside Urinary Proteomics, 2007-2009.

11. Biostatistician (J. Klein, PI, Louisville) 10% effort, Pediatric Clinical Proteomics Center, Department of Energy, 2005-2008.

12. Co-I (P. Epstein, PI, Louisville) 10% effort, NIH R01, Podocytes and Oxidative stress in

4

diabetic Kidney, 2006-2007.

13. Co-PI (K. B. Grant, PI) Brains and Behavior Seed Grant, GSU, $25,414, 2005-2006.

14. Co-PI (I. Weber, PI) Research Program Enhancement Award, GSU, student support, $36000, 2004-2007.

15. Investigator, Student & Travel support for five years, $75000, Georgia Cancer Coalition (Michael Eriksen PI), 2004-2009.

16. Statistician (2 months of summer salary), BimCore, Emory University, summer 2004.

17. Co PI (M. Brinton, PI) Biomedical Computing Center Seed Grant, GSU, $13467, summer 2004.

18. Consultant on a NSF funded project in Structural Biology (B. C. Wang, PI), University of Georgia, $8087, summer 2003.

19. Co-PI (J. Arnold, PI) Genomics and Computational Biology: A REU Site, National Science Foundation, Joint Program between UGA, GA State and Clarke Atlanta University, $210,000, 2003-2005.

20. Co-PI (G. Chen, PI) Tech Fee Grant , GSU, $58002, 2003.

21. Co PI (D. Vidacovic, PI) Instructional Improvement Grant, GSU Center for Teaching & Learning, $5000, 2000-2001.

22. Co PI (E. Dubinsky, PI) IPCURT Project Course and Curriculum Development, National Science Foundation, $100,000, 1998 - 1999.

Book Editing:

Datta, S. and Martens, B. (2016) Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry, Springer, 2016-12-16. ISBN **97833194580**

PUBLICATIONS

Six Most Cited Papers
(Google Scholar, accessed on 02/26/2019)

1. **Datta, S.** and Datta, S. *Bioinformatics*, (2003), Citation Count = 420

2. Brock, G., Pihur, V., **Datta, S.** and Datta, S. (2008). *Journal of Statistical Software*, **25**, 4 Citation Count = 405

3. Pihur, V. Datta, S., Datta, S. (2009) *BMC bioinformatics* Citation Count = 214

4. **Datta, S.**, Datta, S. (2006) *BMC Bioinformatics*, Citation Count = 177

5. Gerrit J. Schut, G., J., Brehm, S., D., **Datta, S**. and Adams, M. W., *Journal of Bacteriology*, July (2003), Citation Count = 166

Refereed papers, refereed book chapters/entries and **newsletter articles:**

2019

1. Mishra, S.; Crowley, P.; Wright, K.; Palmer, S.; Walker, A.; Datta, S.; Brady, L. J. (2019) Membrane proteomic analysis reveals overlapping and independent functions of *Streptococcus mutans* Ffh, YidC1, and YidC2. Accepted in *Molecular Oral Microbiology*.

2. Sekula, M., Gaskins, J., Datta, S. (2018) Detection of differentially expressed genes in discrete single cell RNA. Accepted in *Biometrics*.

3. Walker, A. R., Datta, S. (2018) Identification of city specific important bacterial signature for the MetaSUB CAMDA challenge microbiome data. Accepted in *Biology Direct*.

4. Park, S., Porter, M., Park, Ki, Bielick, L., Rooks, B., Mainous, A., G., Datta, S., Carek, P., J. (2018) Medical Schools, Primary Care, and Family Medicine:Clerkship Directors[1] Perceptions of the Current Environment. Accepted in *Family Medicine*.

5. Wu, Y., Datta, S., Little, B., Kong, M. (2018). Causal Analysis of Dietary Components and Physical Activity in Type 2 Diabetes by Gender in White, African American and Mexican American: NHANES 2011-2014 (submitted).

6. Tadros, H. J. et al. (2019). Meta-analysis of cardiomyopathy-associated mutations in troponin complex-encoding genes identifies hotspots that predispose to worse outcomes (submitted)

(2018)

7. Walker, A. R., Grimes, T. L., Datta, S. **Datta, S.** (2018) Unraveling bacterial fingerprints of city subways from microbiome 16S gene profiles. ***Biology Direct*** 201813:10

8. Grimes, T., Walker, A. R., Datta, S. Datta S. (2018). Predicting survival times for neuroblastoma patients using RNA-seq expression profiles. ***Biology Direct*** 201813:11

9. Mikuls, T. R., Walker, C., Qiu, F., Yu, F., Thile, G. M., Alfant, B., Li, E., Zhao, L., Y., Wang, G., P., **Datta, S.**, Payne, J., B., (2018) The Subgingival Microbiome in Patients with Established Rheumatoid Arthritis. ***Rheumatology***, Volume 57, Issue 7, 1 July 2018, Pages 1162–1172. PMID: 29562298

10. Lyon DE, Starkweather A, Yao Y, Garrett T, Kelly DL, Menzies V, Derezinski P, **Datta S**, Kumar S, Jackson-Cook C. (2018) Pilot Study of Metabolomics and Psychoneurological Symptoms in Women With Early Stage Breast Cancer. ***Biol Res Nurs.*** 2018 Mar; 20(2):227-236.

11. Mitra, R., Gill, R., Sikdar, S., Datta, S. Bayesian Hierarchical Model for Protein Identifications. ***Journal of Applied Statistics***. Published online: 25 Mar 2018

2017

12. Wu, Y*, Gaskins, J., Kong, M., Datta, S. (2017) Profiling the Effects of Short Time-Course Cold Ishemia on Tumor Protein Phosphorylation Using a Bayesian Approach. *Biometrics.* 10.1111/biom.12742

13. Wan, Y., Datta, S., Kong, M. (2017) Monotonic Single-index Models with Application to Assessing Drug Interaction. Feb 20 2017 In : *Statistics in Medicine*. 36, 4, p. 655-670. PMC 5217167

14. Pesonen, M.*, Nevalainen, J., Potter, S. S, Datta, S., **Datta, S**. (2017) A Combined PLS and negative binomial regression model for inferring association networks from next-generation sequencing count data. *IEEE/ACM Trans Comput Biol Bioinform.* 2017 Feb 7. doi: 10.1109/TCBB.2017.2665495. PMID: 28186904. PMCID: PMC5547023.

15. Sikdar, S. and Datta, S. (2017). A novel statistical approach for identification of the master regulator transcription factor. Accepted in *BMC Bioinformatics*. 2 February 2017

16. Sekula, M.*, Datta, S. and Datta, S. (2017) optCluster: An R package for determining the optimal clustering algorithm and number of clusters. *Bioinformation* 13(3): 100-103. PMC545052

17. Dutta, S.*, **Datta, S.** and Datta, S. (2017) Temporal prediction of future state occupation in a multistate model from high dimensional baseline covariates via pseudo-value regression. *Journal of Statistical Computation and Simulation*, Volume 87, 2017 - Issue 7

18. Gill., R., Datta, S. (2017). "Probabilistic and likelihood-based methods for protein identification from MS/MS data", in Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry in Frontiers in Probability and the Statistical Sciences series, Springer

19. Mertens, B. J. A., Datta, S., Hankemeier, T., Beekman, M., Uh, Hae-W. (2017). Statistical analysis of lipidomics data in a case-control study, in Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry in Frontiers in Probability and the Statistical Sciences series, Springer

20. Datta, S. (2017). Advancing Omics Data Analysis: A Call for Participation by a Statistician in the Field, *Chance* 30:2

21. Sikdar, S., Datta, S., and Datta, S. (2017) EAMA: Empirically adjusted meta-analysis for large-scale simultaneous hypothesis testing in genomic experiments. Accepted in PlosOne.

22. Datta, S. (2017) Let's Chat: Top WSDS Speakers Provide Insights Into Their Careers. AmStat News 1 June, 2017.

23. Chakraborty, S., Datta, S., Datta, S. (2017). Disease Modeling and Publuc Health (Special Issue) in HANDBOOK OF STATISTICS Edited Volume by Elsevier. "Nonparametric Regression of State Occupation Probabilities in a Multistate Model" Accepted.

24. Lyon, D., Starkweather, A, Yao, Y., Garrett, T., Lynch-Kelly, D., Menzies, V., Derezinski, P., Datta, S., Kuman, S., Jackson-Cook, C. (2017). Pilot study of the association of global and targeted metabolomics with psychoneurological symptoms in women with early stage breast cancer prior to and after chemotherapy. Accepted in *Biological Research for Nursing.*

25. Mikuls, T., Walker, C., Qiu, F., Fang, Q., Thiele, G., Alfant, B., Li, Eric, Zhao, L., Wang, G., **Datta, S.**, Payne, J. (2017). The Subgingival Microbiome in Patients with Established Rheumatoid Arthritis. Submitted.

2016

26. Siriwardhana, C.*, **Datta, S.**, and Datta, S. (2016). Inter-platform concordance of gene expression data for the prediction of chemical mode of action. ***Biology Direct***, 2016 11:67.

27. Sikdar, S.*, Datta, S., and **Datta, S.** (2016). Exploring the importance of cancer pathways by meta-analysis of differential protein expression networks in three different cancers. ***Biology Direct***, 2016 **11**:65.

2015

28. Datta., S., Lorenz, D., and **Datta, S.** Approximate U-statistics for state waiting times under right censoring. In Modern Multivariate and Robust Methods (K. Nordhausen, S. Taskinen, Eds.), Springer, pp. 31-46 (2015).

29. Sikdar, S., Wosoba, C., H., Abdia, Y., Dutta, S., Gill, R., Datta, S., **Datta, S.** (2015). An Integrative Exploratory Analysis of -omics Data from the ICGC Cancer Genomes Lung Adenocarcinoma Study. ***Systems Biomedicine*** 2:3, 1-9; April 1, 2015.

30. Kujala M, Nevalainen J, März W, Laaksonen R, Datta S (2015) Differential Network Analysis with Multiply Imputed Lipidomic Data. ***PLoS ONE*** 10(3): e0121449.

31. Datta, S. and Gill, R. (2015). Statistical Analysis of DNA Microarray Data. Wiley StatsRef: Statistics Reference Online. 1–7. Update based on original article by Susmita Datta, ***Wiley StatsRef: Statistics Reference Online © 2014 John Wiley & Sons, Ltd.***

32. Sikdar, S., Gill, R., Datta, S. (2015) Protein identification from tandem mass spectrometry data. Invited Paper in ***Briefings in Bioinformatics***, July 3, 2015

2014

33. Gill, R., Datta, S., **Datta, S.** (2014) Differential Network Analysis in Human Cancer Research. ***Current Pharmaceutical Design*** **(Special Issue on S*ystems Biology)**, 20(1):4-10. **PMC4583148**

34. Chakraborty, S.*, Datta, S. and Datta, S. Nonparametric regression of state occupation probabilities in a multistate model via generalized additive models for transition count and number at risk processes. Submitted.

35. Lorenz, D., J., Gill, R. S., Mitra, R., **Datta, S.** (2014) Using RNA-Seq data to detect differentially expressed genes. Accepted in *Statistical Analysis of Next Generation Sequencing Data* (S. Datta and D. Nettleton, Eds.), Springer. ISBN: 978-3-319-07211-1 (Print) 978-3-319-07212-8 (Online). Pages 25-49.

36. Mitra, R., Gill, R. S., **Datta, S.**, Datta, S. Statistical analyses of next generation sequencing data: an overview. Accepted in *Statistical Analysis of Next Generation Sequencing Data* (S. Datta and D. Nettleton, Eds.), Springer. ISBN: 978-3-319-07211-1 (Print) 978-3-319-07212-8 (Online). Pages 1-24.

37. Wan, Y., **Datta, S.**, Conklin, D.J., Kong, M. (2014) "Variable Selection Models Based on Multiple Imputation with an Application for Predicting Median Effective Dose and Maximum Effect", *Journal of Statistical Computation and Simulation* 4/14/14, DOI: 10.1080/00949655.2014.90780.

38. Gill, R., Datta, S., Datta, S. (2014) dna: An R package for differential network analysis. Bioinformation. 2014; 10(4): 233–234. PMCID: PMC4070055

39. Gill, R., Datta, S., **Datta, S.** (2014) Bridging in vivo and in vitro data from Japanese Toxicogenomics Project using network analyses. *Systems Biomedicine*. Published online: 31 Oct 2014. DOI: http://dx.doi.org/10.4161/sysb.28527

40. Shah, J. S.‡, Datta, S. and Datta, S. A (2014) multi-loss super regression learner via bagging and rank aggregation with application to survival prediction using proteomics. *Computational Statistics* 29(6): 1749-67.

41. Webb-Robertson B-J., Matzke, M. M., **Datta, S.**, Payne, S. H., Nicora, C. D., Shukla, A. K., Metz, T. O., Rodland, K. D., Smith, R. D., Tardiff, M. F., Waters, K. M., McDermott, J. E., Pounds J. G. and (2013) Bayesian Proteoform Modeling Improves Protein Quantification of Global Proteomic Experiments. Accepted in *Molecular and Cellular Proteomics.* Aug 18, 2014. PMID 25129695

42. Litchfield , L. M., Appana, S. N., **Datta, S.**, Klinge, C. M. (2014) *COUP-TFII inhibits NFkB activation in breast cancer cells and restores endocrine sensitivity,* Molecular and Cellular Endocrinology 382 (2014) 358-367.

43. Gill, R., Datta, S. and Datta, S. (2014). Differential Network Analysis in Human Cancer, Current Pharmaceutical Design Volume 20, Issue 1, 4-10 (7).

44. Datta, S. (2014). Letter from the Past President, *Newsletter for Caucus for Women in Statistics*, Volume 44, No 1: Fall, 2014.

2013

45. Chakraborty, S., **Datta, S.** and Datta, S. (2013). Svapls: An R package to identify and correct for hidden variation in gene expression data. *BMC Bioinformatics* 2013, **14**:236 Software. **Highly Accessed**

46. Mostajabi, F., Datta, S., Datta, S. (2013) Predicting Patient Survival from Proteomic Profiles using Mass Spectrometry Data: An Empirical Study. *Communications in Statistics - Simulation and Computation*, 42, 485-498.

47. Datta, S. (2013). Letter from President, *Newsletter for Caucus for Women in Statistics*, Volume 43, No 1: Winter, 2013.

48. Datta, S. (2013). Letter from President, *Newsletter for Caucus for Women in Statistics*, Volume 43, No 2: Spring, 2013.

49. Datta, S. (2013). Letter from President, *Newsletter for Caucus for Women in Statistics*, Volume 43, No 3: Fall, 2013.

50. Bible, J., **Datta, S.** and Datta, S. (2013). Cluster analysis: Finding groups in data. In *Informatics for Materials Science and Engineering* (K. Rajan, Ed.) Elsevier, pp 53-70.

51. Gill, R., Datta, S. and Datta, S. Similarity in network structures for *in vivo* and *in vitro*

data from the Japanese Toxicogenomics Project. In *CAMDA 2013 Proceedings*, online @http://dokuwiki.bioinf.jku.at/lib/exe/fetch.php/gill.pdf **(2013)**.

52. Datta S., (2013) Feature selection and machine learning with mass spectrometry data. *Methods Mol Biol.*, Springer. 2013;1007:237-62. doi: 10.1007/978-1-62703-392-3_10.

2012

53. Chakraborty, S., Datta, S. and **Datta, S.** (2012) Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies. *Bioinformatics*, 28 (6): 799-806.

54. Litchfield, L. M., Riggs, K. A., Emberts, C. G., Hockenberry, A. M., McConda, D. B, Oliver, L. D., Fox, J. M., Hu, C., Cai, J., Pierce, W. M., Jr., Ivanova, M. M., Bates, P. J., Martin, R. G. C., Appana, S. N., **Datta, S**., Kulesza⸴ P., and Klinge, C. M. Regulation of COUP-TFII transcriptional activity by interaction with nucleolin in human breast cancer cells and tumors (2012). *PLoS ONE*, 7(5):e38278. Epub 2012 May 31.

55. Li, X., Klinge, C. M., Datta, S. Novel and Alternative Bioinformatics Approaches to Understand miRNA-mRNA Interactome in Cancer Research. In *Systems Biology in Cancer Research and Drug Discovery* (Azmi, A. Ed) Springer, 267-288.

56. Datta, S. (2012). Letter from President-Elect, *NewsLetter for Caucus for Women in Statistics*, Winter 2012.

57. Datta, S. (2012). Letter from President-Elect, *NewsLetter for Caucus for Women in Statistics*, Fall 2012.

2011

58. Ndukum, J., Fonseca, L, L., Santos, H., Voit, E., O. And **Datta, S.** (2011). Statistical Inference Methods for Sparse Biological Time Series Data, *BMC Systems Biology*, 5**,** 57. PMC3114728

59. Li, X, Gill, R., Cooper, N., G., F., Yoo, J., K., and **Datta, S.** (2011). Modeling microRNA-mRNA Interactions Using PLS Regression in Human Colon Cancer, *BMC Medical Genomics*, **4**, 44. PMC3123543 **Highly Accessed**

60. Ndukum, J., Atlas, M., **Datta, S.** (2011). pkDACLASS: open source software for analyzing MALDI-TOF, *Bioinformation*, **6**, 45-47. PMC3064853

61. Manavalan, T. T., Teng, Y., Bhimani, S., Appana, S. N., **Datta, S**., Kalbfleisch, T. S., Li, Y., and Klinge, C. M. (2011). Differential expression of microRNA expression in tamoxifen-sensitive MCF-7 versus tamoxifen-resistant LY2 human breast cancer cells, *Cancer Letters*, **313**, 26-43.

62. Pihur, V., Datta, S., **Datta, S.** (2011). Meta analysis of Chronic Fatigue Syndrome through integration of clinical, gene expression, SNP and proteomic data, *Bioinformation*, **6**, 120-124. PMC3089886

2010

63. **Datta, S,** Pihur, V. and Datta, S. (2010). An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data, *BMC Bioinformatics*, **11**, 427.

64. Gill, R., Datta, S. and **Datta, S.** (2010). A statistical framework for differential network analysis from microarray data using partial least squares, *BMC Bioinformatics*, **11**, 95.

10

PMC2838870 **Highly accessed**

65. **Datta, S**., Datta, S., Kim, S., Chakraborty, S. and Gill, R. S Statistical Analyses of Next Generation Sequence Data: A Partial Overview (2010). *Journal of Proteomics & Bioinformatics*, **3**, 183-190. PMC2989618.

2009

66. Pihur, V., Brock, G., **Datta, S.** and Datta, S. Cluster validation for microarray data: An appraisal. (2009). In *Multivariate Statistical Methods*, (A. Sengupta, Ed), ISI Platinum Jubilee series, Vol 5, World Scientific Press, pp. 79-94.

67. Datta, S. and Datta, S. (2009). Computational biology touches all bases. *Genome Biology*, 10, 303.

68. Pihur, V., **Datta, S.** and Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. *BMC Bioinformatics*, **10**, 62. **Highly Accessed**

69. Yoo, J. K., Becky S. Patterson, B. S. and **Datta, S.** (2009). OLS-based predictor test in single index model to predict transcription rate by histone acetylation level, *Statistics & Probability Letters*, **79**, 2109-2114.

70. Atlas, M. and **Datta, S.** Monoisotopic Peak Detection for Mass spectrometry Data (2009). *Journal of Proteomics and Bioinformatics*, **2**, 202-216.

71. **Datta, S.** and Pihur, V. (2009). Feature selection and machine learning with mass spectrometry data, R. Matthiesen, ed., In *Clinical Proteomics: Methods, Applications and Tools, Humana Press*, (Matthiesen, R. ed.), pp. 205-229.

2008

72. Brock, G., Pihur, V., **Datta, S.** and Datta, S. (2008). clValid, an R package for cluster validation. *Journal of Statistical Software*, **25**, 4.

73. Pihur, V., Datta, S. and **Datta, S.** (2008). Finding cancer genes through meta-analysis of microarray experiments: Rank aggregation via the cross entropy algorithm. *Genomics*, **92**, 400-403.

74. Pihur, V., Datta, S. and **Datta, S.** (2008). Reconstruction of genetic association networks from microarray data: A partial least squares approach. *Bioinformatics*, **24**, 561-568.

75. **Datta, S.**, Turner, D., Singh, R., Ruset, B., Pierce, W. M., and Knudsen, T. B. (2008). Fetal alcohol syndrome in mice detected through proteomics screening of the amniotic fluid. *Birth Defects Research Part A: Clinical and Molecular Teratology*, **82**, 177-186.

2007

76. **Datta, S.**, Le-Rademacher, J. and Datta, S. (2007). Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO, *Biometrics*, **63**, 259-271.

77. **Datta, S.**, Datta, S., Parrish, R. S. and Thompson, C. M. (2007). Microarray data analysis. In *Computational Methods in Biomedical Research*, R. Khatree and D. Naik, eds., Chapman & Hall/CRC Biostatistics Series, Volume 24, 1-43.

78. Boratyn, G. M., **Datta, S.** and Datta, S. (2007). Incorporation of biological knowledge into distance for clustering genes. *Bioinformation*, **1**, 396-405.

79. Pihur, V., **Datta, S.** and Datta, S. (2007). Weighted rank aggregation of cluster validation measures: A Monte Carlo cross-entropy approach. *Bioinformatics*, **23**, 1607-1615.

80. Pihur, V., **Datta, S.** and Datta, S. (2007). Understanding Chronic Fatigue Syndrome (CFS) from CAMDA data: A systems biology approach. In *Proceedings of CAMDA 2007*, full paper, online @ http://camda.bioinfo.cipf.es/camda07/agenda/detailed.html.

2006

81. **Datta, S.** and de Padilla, L.M. (2006). Feature selection and machine learning with mass spectrometry data for distinguishing cancer and non-cancer samples, *Statistical Methodology* (Special Issue on *Bioinformatics*), **3**, 79-92.

82. **Datta, S.** and Datta, S. (2006). Validation measures for clustering algorithms incorporating biological information, *IEEE Proceedings of International Multi-Symposiums on Computer and Computational Sciences* **(IMSCCS|06)**, (J. Ni, J. Dongarra, Y. Zheng, G. Gu, G. Wolfgang and H. Jin, Eds.), **1**, 131-135.

83. **Datta, S.** and Datta, S. (2006). Evaluation of clustering algorithms for gene expression data, *BMC Bioinformatics*, **7** (Suppl 4), S17.

84. **Datta, S.** and Datta, S. (2006). Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes, *BMC Bioinformatics*, **7**, 397. (**Highly Accessed**)

85. Boratyn, G. M., Datta**, S.** and Datta, S. (2006). Biologically supervised hierarchical clustering algorithms for gene expression data, In *Proceedings of the 28th IEEE EMBS Annual International Conference,* New York City, USA, 5515-5518.

2005

86. **Datta, S.** and Datta, S. (2005). Empirical Bayes screening (EBS) of many p-values with applications to microarray studies, *Bioinformatics*, **21**, 1987-1994.

87. Weinberg, M. V., Schut, G. J., Brehm, S., Datta, S., and Adams, M. W. W. (2005). A hyperthermoplilic cold shock response: the archaeon Pyrococcus furiosus synthesizes novel membrane-bound glycoproteins at a sub-optimal growth temperature. *Journal of Bacteriology*, 187, 336-348.

88. Datta, S. (2005). Statistics in Genetics, In *Encyclopedia of Statistical Sciences*, Second edition, Wiley, New York.

89. Datta, S. (2005). Statistics in Microarray Analysis, In *Encyclopedia of Statistical Sciences*, Second edition, Wiley, New York.

90. Datta, S. (2005). Statistics in Vaccine Studies, In *Encyclopedia of Statistical Sciences*, Second edition, Wiley, New York.

2004

91. Datta, S., Satten, G. A., Benos, D. J., Xia, J., Heslin, M., and Datta, S. (2004). An empirical Bayes adjustment to increase the sensitivity of detecting differentially expressed genes in microarray experiments, *Bioinformatics*, 20, 235-242.

92. **Datta, S.** and Datta, S. (2004). An empirical Bayes adjustment to multiple p-values for the detection of differentially expressed genes in microarray experiments. In *Bioinformatics 2004*, *Conferences in Research and Practice in Information Technology – Second Asia-Pacific Bioinformatics Conference*, **29**, Y-P. P. Chen, Ed., 155-159, Australian Computer Society, Sydney.

93. Warrenfeltz, Z., Pavlik, S., Datta, S., Kraemer, E., Benedict, B. Mcdonald, J. F. (2004). Gene expression profiling of epithelial ovarian tumors correlated with malignant potential. *Molecular Cancer*, 3, 27.

2003

94. Datta, S. (2003). Statistical techniques for microarray data: A partial overview, *Communications in Statistics-Theory and Methods*, **32**, 263-280.

95. **Datta, S.** and Datta, S. (2003) Comparisons and validation of statistical clustering techniques for microarray gene expression data*, Bioinformatics*, **19**, 459-466.

96. Arnold, J., Schuttler, H.-B.,Logan, D., Griffith, J., Arpinar, B. **Datta, S.**, Kochut, K. J., Kraemer, E., Miller, J. A., Sheth, A., Aleman-Meza, B., Doss, J., Harris, L. and Nyong, A. (2003). Metabolomics, In *Handbook of Industrial Mycology, Chapter 23*. Marcel-Dekker, New York, NY.

97. Schut, J., G., Brehm, S., **Datta, S.**, and Adams, M. W. W. (2003). Whole Genome DNA microarray of a hyperthermophile and an archaeon: Pyrococcus furious grown on peptides and carbohydrate, *Journal of Bacteriology*, **185**, 3935-3947.

2002

98. **Datta, S.** and Arnold, J. (2002). Some comparisons of clustering and classification techniques applied to transcriptional profiling data. In *Advances in Statistics, Combinatorics and Related Areas,* Eds.: C. Gulati, Y-X. Lin, S. Mishra, and J. Rayner, World Scientific, 63-74.

2001

99. Datta, S. (2001). Estimation of selection parameters using multi-generation cytonuclear data, *Biometrical Journal*, **43**, 219-233.

100. Datta, S. (2001). Exploring relationships in gene expressions: A partial least squares approach, *Gene Expression*, **9**, 257-264.

101. Datta, S. (2001). Testing neutrality of mtDNA using multigeneration cytonuclear data, Selected Proceedings of the Symposium on Inference for Stochastic Processes, Eds.: I. V. Basawa, C. C. Heyde and R. L. Taylor, *IMS Lecture Notes - Monograph Series, 37, 173- 184, IMS, Beachwood, OH*.

2000

102.   Datta, S. (2000). Some statistical aspects of cytonuclear disequilibria. In **Statistics in Molecular Biology and Genetics**, Ed: Francoise Seillier-Moiseiwitsch, IMS Lecture Notes- Monograph Series, **33**, 21-37.

103.   **Datta, S.**, Satten, G. A. and Datta, S. (2000). Nonparametric estimation for the three stage irreversible illness-death model, **Biometrics**, **56**, 841-847.

104.   Datta, S. (2000). Some statistical issues involving multi-generation cytonuclear data, In **Advances on Methodological and Applied Aspects of Probability and Statistics**, N. Balakrishnan, Ed., Gordon and Breach, 525-546.

105.   **Datta, S.**, Satten, G. A. and Datta, S. (2000). Estimation of stage occupation probabilities in multistage models, In **Advances on Theoretical and Methodological Aspects of Probability and Statistics**, N. Balakrishnan, Ed., Gordon and Breach, 493-506.

106.   Datta, S. (2000). Book Review: Statistics in Human Genetics by Pak Sham. **Statistics in Medicine**, **19**, 1384-1385.

1999

107.   Scribner, K. T., **Datta, S.**, Arnold, J., and Avise, J. C. (1999). Empirical evaluation of cytonuclear models incorporating genetic drift and tests for neutrality of mtDNA variants: data from experimental Gambusia hybrid zones, **Genetica**, **105**, 101-108.

108.   **Datta, S.**, Halloran, E. M. and Longini, I. M. (1999). Efficiency of estimating vaccine efficacy for susceptibility and infectiousness: randomization by individual versus household, **Biometrics**, **55**, 792-798.

1998

109.   **Datta, S.** and Arnold, J. (1998). Dynamics of cytonuclear disequilibria in subdivided populations, **Journal of Theoretical Biology**, **192**, 99-111.

1997

110.   **Datta, S.**, Longini, I. M., and Halloran, E. (1997). Measuring vaccine efficacy for different HIV vaccine trials, **Statistics in Medicine**, **17**, 185-200.

1996

111.   **Datta, S.**, Fu, Y. X., Arnold, J. (1996). Dynamics and equilibrium behavior of cytonuclear disequilibria under genetic drift, mutation, and migration, **Theoretical Population Biology**, **50**, 298-324.

112.   **Datta, S.** and Arnold, J. (1996). Diagnostics and a statistical test of neutrality hypothesis using the dynamics of cytonuclear disequilibria, **Biometrics**, **52**, 1042-1054.

113.   **Datta, S.**, Rand, D. M., and Arnold, J. (1996). A statistical test of a neutral model using the dynamics of cytonuclear disequilibria, **Genetics**, **144**, 1985-1992.

114.   Longini, I. M., **Datta, S.**, and Halloran, E. (1996). Measuring vaccine efficacy for both susceptibility to infection in infectiousness for prophylactic HIV-1 vaccines, **Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology,** **13**, 440-447.

Non-Refereed research papers:

2005

115.　　　**Datta, S.** and Datta, S. (2006). Validation of statistical clustering using biological information, ***Proceedings of INTERFACE 2005*** (CD-ROM).

1999

116.　　　Datta, S. (1999). Hypotheses testing for different selection models using multi-generation cytonuclear data, ***Proceedings of American Statistical Association, Biometrics Section***, 157-161, Alexandria, USA.

EDITORIAL AND REFERRING SERVICE:

2018-　　　One of the four Book Series Editors for the Springer-IISA Series on Statistics and Data Science (2018-

2016　　　Co-Editor of the book on "Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry in Frontiers in Probability and the Statistical Sciences series, Springer, 2016

2013-　　　Section Chief Editor, Bioinformation

2010-　　　Editorial Board Member, Briefings in Bioinformatics (Impact Factor 9.283)

2008-　　　Associate Editor, BMC Research Notes

2007-　　　Associate Editor, Bioinformation

2006　　　Editorial Board Member, Bioinformation

2007-　　　Special Issue Editor (Gene Expression Analysis), Bioinformation, Vol. 1, Issue 10.

2007-2011　Associate Editor, Statistical Methodology

2007-2013　Associate Editor, Statistics & Probability Letters

Ad-hoc Member/Reviewer:

2019　　　Informatics Technologies for Cancer Research and Surveillance (ITCRS), U01/U24/UG3/R21 grant review panel, March 21 and 22, 2019

2018　　　ZGM1 TRN-X CF, Review of R13's Conference Grants, Dec 7, 2018

2018　　　RFA-RM-17-012 "Metabolomic Data Analysis and Interpretation Tools (U01)" Study section review panel. March 23, 2018

2017　　　ZRG1 PSE-H 07 - Infectious, Reproductive, Asthma and Pulmonary Conditions: Additional Applications. (Mail Reviewer)

2016　　　ZRG1 BST-U(50), R03 Metabolomics Review Panel, NIH, June 22-23, 2016

2016　　　Global Omics Review Panel , NIH, January, 2016

2015　　　Modeling and Analysis of Biological Systems Study Section (MABS) June 11-12, 2015

2015　　　National Institute of Health, Special Emphasis Panel for P01 proposal ZRG1 BST-N (40) March 30, 2015

2014　　　National Institute of Health/BMRD study section reviewer for PA13-304

2014　　　National Institute of Health/NRCS ad hoc study section reviewer for PA13-264

2014　　　National Institute of Health/NIAID ad hoc study section ZAI1 RRS-M (J1) reviewer of investigator initiated program project PAR-13-254(P01) proposal

2014　　　Gertrude M. Cox student award judge on behalf of Caucus for Women in Statistics.

2014　　　Mail Reviewer for Mitacs (Canada) Accelerate research proposal, April, 2014

2014　　　CAMDA 2014 Conference Scientific Committee, Boston, USA.

2014	Co-Chair poster session on "Statistical Significance" on behalf of Scientific and Public Affairs Committee of ASA at JSM, Boston 2014.
2013	Mail Reviewer for National Institute of Health NIH Director's Early Independence Award (DP5), ZRG1 BBBP-E 53 R
2012	National Institute of Health/NIEHS review of the NIEHS Biostatistics Branch
2012	National Institute of Health/NCI, ZCA1SRLB2J1 Cancer Prevention Research R03 Study Section
2012	National Institute of Health, ZRG1 BST-H (50) Special Emphasis Study Section on Bioengineering Sciences and Technologies for U01 metabolomics RFA.
2012	National Institute of Health, ZRG1 BST-M (02) Special Emphasis Study Section of Topics in Bioengineering Sciences and Technologies.
2011	Book proposal review (2011) Title: Advances in Statistical Bioinformatics: Models and Integrative Inference for High-Throughput Data. Publisher: Cambridge University Press.
2008	National Science Foundation, Biology Program, Mail Reviewer, May 2008.
2003	National Institute of Health, Bio-defense Study Section, April 2003.
2003	(Invited) Emtech Bio Scientific Advisory Board members and Seed Grant Reviewers: Georgia Tech, Atlanta, October 2002.
2002	(Invited) Emtech Bio Scientific Advisory Board members and Seed Grant Reviewers: Georgia Tech, Atlanta, October 2002.
2001	Member, Advisory Panel for MRI Program, National Science Foundation, 2001-2002.
2000-2005	Mathematical Review
2000	Book review for Statistics and Medicine, 2000.
2000-	Paper reviewer of multiple ISCB and CAMDA conferences as program Committee members, several years.

Referee work:

Science Translational Medicine, Biometrics, Biostatistics, Journal of American Statistical association, Journal of Statistical Planning and Inference, Statistics and Its Interface, Statistics in Medicine, Journal of Applied Statistics, Computational Statistics & Data Analysis, Journal of Multivariate Statistics, Scandinavian Journal of Statistics, Communications in Statistics, Statistics and Interface, Computational Statistics and Data Analysis, Statistical Applications in Genetics and Molecular Biology, Mathematical Biosciences, Bioinformatics, BMC Bioinformatics, BMC Research Notes, Briefings in Bioinformatics, Journal of Bioinformatics and Computational Biology, BMC Systems Biology, BMC Health Services Research, Proceedings of National Academy of Sciences, Nucleic Acids Research, Proteomics, Biotechnology, Genomics, Journal of Proteome Research, Human Reproduction, Pattern Recognition, International Journal of Data mining and Bioinformatics, Computational Biology and Bioinformatics, Physica A: Statistical Mechanics and Its Applications, Int. J. Developmental Neuroscience, Computer Methods and Programs in Biomedicine, Chemometrics, Computer Methods and Programs in Biomedicine, Environmental Health Perspectives. Moreover, refereed many manuscripts for the conference submissions to Intelligence Systems for Molecular Biology (ISMB) and European Conference on Computational Biology (ECCB) and Critical Assessment of Massive Data Analysis (CAMDA) as a program committee member **multiple times**.

MEMBERSHIPS AT PROFESSIONAL SOCIETIES:

2018-	Association for Clinical and Translational Science (ACTS)
2011-	Caucus for Women in Statistics (CWIS)
2006-	American Association for the Advancement of Science (AAAS)

2007-   International Statistical Institute (ISI)
2000-   International Biometric Society (ENAR)
1999-   International Society for Computational Biology (ISCB)
1997-   International Indian Statistical Association (Life) (IISA)
1997-   Forum for Interdisciplinary Mathematics (Life) (FCM)
1995-   American Statistical Association (ASA)
1995-   Institute of Mathematical Statistics (IMS)

SERVICE TO STATISTICAL AND COMPUTATIONAL BIOLOGY ORGANIZATIONS

2019   University Institutional Review Board Member
2019   University IT committee
2018-2019 Curriculum Committee Chair of School of Public Health and Health Professions, University of Florida
2018   Program Committee Chair of The SRCOS (Southern regional Council of Statistics) 2018 Summer Research Conference, Virginia Beach, VA
2018   Program Committee Member of the Intelligent Systems for Molecular Biology (ISMB), 2018, July 6-10, 2018, in Chicago, Illinoi.
2018   Program Committee and Local Organizing Committee, Critical assessment of Massive Data Analysis (CAMDA) 2018, July 6-10, 2018, **in Chicago, Illinoi.**
2017   Executive Committee Member of the Women in Statistics and Data Science, October 19-21, 2017.
2016-   Co-Director of Biostatistics Epidemiology Research Design, CTSI, University of Florida
2016-   Chair of the Biostatistics Curriculum Committee, University of Florida
2015   Program Committee Member of the Intelligent Systems for Molecular Biology (ISMB), 2016, July 8-12, 2016, in Orlando, Florida.
2016   Program Committee and Local Organizing Committee, Critical assessment of Massive Data Analysis (CAMDA) 2016, June 2016, Orlando, Florida.
2016   IT – Research Steering Committee, Health Sciences, University of Florida.
2016   Chair: ENAR Webinar Committee
2014   Executive Committee Member of the first conference Celebrating Women in Statistics. Raleigh, NC, May 15-17 2014.
2014-16   Appointed Member of Regional Advisory Committee (RAB), International Biometric Society (ENAR).
2014   Program Committee Member of the 13th CAMDA conference Boston, 2013.
2014   Program Committee Member of the 13th European Conference on Computational Biology, Strasbourg, Sep 7-10, 2014
2014   Program Committee Member of 21st Annual International Conference on Intelligent Systems for Molecular Biology, ISMB, Boston 2014.
2014   Ex-Officio member of ASA committee of Women in Statistics
2011-2014 Appointed member of the Executive Committee for the Conference of Women in Statistics 2014.
2013   **Elected President** of the Caucus for Women in Statistics
2013   Program Committee Member of 21st Annual International Conference on Intelligent Systems for Molecular Biology, ISMB/ECCB 2013, The 12th European Conference on Computational Biology, Berlin, July, 19-23, 2013.
2013   Program Committee Member of The 12th CAMDA conference Berlin, Germany July 19 to July 20, 2013.
2011-2015 Elected member of ASA Scientific & Public Affairs Advisory Committee
2012-2012 Appointed member of the COWIS-Gertrude M. Cox Scholarship Subcommittee of ASA.
2012   CAMDA 2012 Conference Scientific Committee, Long Beach, California.
2012-13   Elected President of Caucus for Women in Statistics
2011-2011 Elected Representative-at-Large Caucus for Women in Statistics

2011   Breakfast Round Table Discussion Coordinator for the Caucus for Women in Statistics, JSM 2011.

2011   Co-Chair of the 'Statistical Significance' Poster Competition organized by the ASA Scientific & Public Affairs Advisory Committee, JSM 2011.

2011   Program Committee Member of 19th Annual International Conference on Intelligent Systems for Molecular Biology ISMB and ECCB10, The 10th European Conference on Computational Biology, Vienna, July, 15-16, 2011.

2010   Program Committee Member ECCB10, the 9th European Conference on Computational Biology, Ghent, Belgium, Sept 26-29, 2010.

2009   CAMDA 2009 Conference Scientific Committee, Chicago, IL, U.S.A., October 5-6, 2009.

2008   CAMDA 2008 Conference Scientific Committee, Vienna, Austria, December 2008.

2008   Program Committee member, Frontiers of Probability and Statistical Science, Connecticut-Storrs, May 2008.

2007   CAMDA 2007 Conference Scientific Committee, Valencia, Spain, December 2007Program Committee Member, ISMB 2007, Vienna, Austria, July 2007.

2005   Program Committee Member, ISMB 2005, Michigan, July 2005.

2003   Executive Board Member and President of Young Professional Statisticians, International Indian Statistical Association IISA.

ORGANIZATION OF CONFERENCE SESSIONS

2013   Organized "Caucus for Women in Statistics Business Meeting and Social" (Dedicated to International Year of Statistics). 8 August, 2013, Montreal, Canada (with three talks)

2008   Invited session organizer at **JSM 2008**, A New Paradigm of Statistical Data Analysis: Omics Data, Denver, August 2008.

2007   Program Committee Member, ISMB 2007, Vienna, Austria, July 2007.

2007   Chair, Invited session at **JSM 2007**, Inference for Multistate Data under Complex Censoring Structures , Salt Lake City, July - August, 2007.

2006   Invited session organizer, Statistics in Genomics and Proteomics, International Biometric Society Conference IBC 2006, Montreal, Canada, July, 2006.

2004   Invited session organizer Statistics in Genomics, **JSM Toronto, August 2004.**

2004   Invited session organizer, Genetic Data Analysis, International Conference on Statistics in Health Sciences, Nantes, France, June 2004.

2002   Co-organizer, student paper competition for the IISA conference, Athens, GA, May 2004.

2003   Organized (and chaired) an invited session titled "Recent Contributions in Bioinformatics" at **JSM San Francisco, August, 2003**.

2003   Executive Board Member and President of Young Professional Statisticians, IISA.

2002   Organized an invited session titled Survival Skills for Young Statisticians at the IIISA International Conference on Statistics, Probability and Related Areas, DeKalb, Illinois, June 2002.

2002   Organized an invited session on Bioinformatics at SCRA 2002-FIM IX: Ninth International Conference of Forum for Interdisciplinary Mathematics on Statistics Combinatorics and Related Areas, Department of Statistics and Department of Mathematics: University of Allahabad, Allahabad, UP 211 002, India, December 21-23, 2002.

2001   Invited Session Organizer, ENAR, 2001 Joint Statistical Meeting, 2001, Atlanta, Georgia.

2001   Organized an invited session on Statistics in Bioinformatics at the International Conference on Statistics, Combinatorics and Related Areas and the Eighth

International Conference of the Forum for Interdisciplinary Mathematics, Wollongong, Australia, December 2001.

2001	Chair, (invited session) Bioinformatics: Statistical Perspectives and Controversies' at International Conference on Statistics, Combinatorics and Related Areas and the Eighth International Conference of the Forum for Interdisciplinary Mathematics December 2001.
2001	Invited Session Organizer, ENAR, 2001 Joint Statistical Meeting, 2001, Atlanta, Georgia.
1999	Session Chair, Statistical Genetics, ENAR Spring Meeting, 1999, Atlanta, Georgia.
1999	Local Organizing Committee, ENAR Spring Meeting, 1999, Atlanta, Georgia.
1999	Session Chair, Applications of State-Space Modeling in the Science, Special Contributed Session, **Joint Statistical Meeting, 1999, Baltimore, Maryland.**

SERVICE TO ACADEMIC INSTITUTIONS

@ University of Florida

Chair of Curriculum Committee, School of Public Health and Health and Health Professions, University of Florida (2018-19)

Chair of Curriculum Committee, Department of Biostatistics

Member of University IT Committee

Member of Clinical Faculty Recruitment Committee, Department of Biostatistics

Member of College (PHHP) Curriculum committee

PhD Preliminary Examination Committee, Department of Biostatistics

@ University of Louisville

2014-	Chair of faculty recruitment committee, Department of Bioinformatics and Biostatistics, University of Louisville
2014-	Chair of Comprehensive Examination Revisions committee, Department of Bioinformatics and Biostatistics, University of Louisville
2013-	ad hoc program review groups with the EVPRI UofL Re: Seminars program
2013-	ad hoc program review groups with the EVPRI UofL Re: Internal grants programs
2013-	Member of the Promotion and Tenure Committee, School of Public Health and Information Sciences, University of Louisville
2012-	**Graduate Director**, Ph.D. and M.S. program, Department of Bioinformatics and Biostatistics, University of Louisville
2012-	Faculty Advisory Committee, Conn Center of Renewable Energy, University of Louisville (university level)
2012-	Graduate Fellow Selection Committee, Conn Center of Renewable Energy, University of Louisville (university level)
2012-	LAC Award for Excellence in Energy Research Nomination and Selection Committee, Conn Center of Renewable Energy, University of Louisville (university level)
2011-	Bioinformatics Executive Committee member (university level), University of Louisville

2010-   Primary member of University Faculty Grievance Committee

2008   Chair of Bioinformatics Faculty Recruitment Committee

2007   Member, Biostatics Director Search Committee of Brown Cancer Center

2006-2008   Member, Biostatistics Faculty Recruitment Committees

2006-2011   Biostatistics Group Leader for the Center for Genomics and Integrated
            Biology, University of Louisville

2006-11   Biostatistics Group Leader for the Alcohol Center at University of Louisville

2005-2009   Member of Bioinformatics Task force (university level), University of Louisville

2005-   Chair of Graduate Admissions Committee (department level)

2006-   Bioinformatics PhD Program Development Committee (university level)

2006-   Chair of Bioinformatics Emphasis development for the PhD program in
            Biostatistics (department level).

@ Georgia State University

2000-2005   Bioinformatics Executive Committee Member, School of Arts and Sciences
            GA State University
1998-2004   Member of Faculty Recruitment Committees
1998   Chair of Biostatistics Program Development Committee
1999   Chair of Bioinformatics Program Development Committee

External Reviewer for Tenure and Promotion Evaluation

2018   Department of Biostatistics, University of Pittsburg
2017   Department of Biostatistics, University of Pittsburg
2014   Sponsor ISI fellowship
2012   National Institute of Environmental Health Sciences (NIEHS/NIH)
2010   Department of Statistics, Purdue University
2009   Department of Statistics, University of Missouri, Columbia
2007   Department of Biostatistics, Medical College of Georgia

Short Academic Visits

2016   Department of Probability and Statistics, National University of Uzbekistan,
        Tashkent, Uzbekistan, May 2016.
2013   Department of Statistics, Tunghai University, Taichung, Taiwan, December 2013.
2013   Institute of Mathematical Sciences, Faculty of Science, University of Malaya,
        KualaLumpur, Malaysia, June 2013.
2011   School of Public Health, University of Tampere, June 2011.
2010   Department of Statistics, Southwest Jiaotong University, Chengdu,
        China, June, 2010.
2004   Department of Statistics and Applied Probability, National University
        of Singapore, December 2004.
1999   Math Stat Division, Indian Statistical Institute, Calcutta, July 1999.

HONORS AND AWARDS:

2016	Nominated as a member-at-large American Association of Advancement of Sciences (AAAS).
2015-2020	Selected member of Committee of Women in Statistics, American Statistical Association (ASA).
2014	Featured on Statistics Views by Wiley "I really think nobody has any choice but to embrace numbers": Susmita Datta looks back on her Presidency of CWIS
2014 -	Elected Fellow of American Association of Advancement of Sciences (AAAS)
2014-16	Appointed Member of Regional Advisory Committee, International Biometric Society (ENAR).
2013	Appointed as a Distinguished University Scholar for outstanding research, scholarship and creative activities
2013	Elected President for the Caucus for Women in Statistics
2012	Elected Fellow of American Statistical Association (ASA) for outstanding contributions to the statistical profession
2011-2015	Elected member of ASA Scientific and Public Affairs Advisory Committee
2010-2013	Elected Representative-at-large for Women in Statistics
2010 (July)	Appointed as a University Scholar for outstanding research, scholarship and creative activities
2010	Biography inclusion in 2010 (64th Edition) of Marquis Who's Who in America and then in Who's Who in the world, in Science and Engineering and Who's Who in women
2007	Elected member of International Statistical Institute, October 2007
2007	Nominated for Provost's Award for exemplary advising, University of Louisville
2003	Appeared in Fox News Atlanta, Comment on State Lottery System
2003	Featured Research faculty in College of Arts and Sciences, Georgia State University, Feb., 2003.
2002	Co-recipient of the CURO Excellence in Undergraduate Research Mentoring Award from University of Georgia, April 2002.
2001-2003	Executive Board Member and Elected President of Young Professional Statisticians, International Indian Statistical Association (IISA) 2001-2003
2002	Press coverage (Atlanta Business Chronicle), April, 2002
2002	Press coverage (Georgia State Magazine), Fall, 2002
2001	NCI travel award for "Workshops for Junior Biostatisticians, 2001 ENAR", Charlotte, N. Carolina.
2000	Outstanding Junior Faculty Award nomination, Georgia State University, Atlanta, Georgia, April 2000
2000	Phi Kappa Phi honor society, Georgia State University Chapter, acceptance, April 2000
1998	NSF Travel Award for IBC98, Cape Town, South Africa, December 1998.
1998	NSF Travel Award for Pathways to the Future workshop, Dallas, Texas, August 1998
1995	Student paper award in SRCOS/ASA summer conference, Melbourne, Florida, June 1995
1994	Best Theoretical Student Award, Department of Statistics, University of Georgia, Athens, Georgia, 1994

INVITED PRESENTATIONS AT NATIONAL OR INTERNATIONAL CONFERENCES

2019 "Advances and Challenges in Single Cell RNA-Seq Analysis", Key note presentation at the Kansas-Western Missouri Chapter of the ASA, Manhattan, Kansas, April 6, 2019

2019   "

2017 "Reverse engineering of genomic networks by statistical association scores", Special Invited
     Talk, IISA, Dec 28, 2017, Hyderabad, India

2017 "Are Statisticians Prepared for the Data Science Challenge? - A Career Development Panel",
     IISA, Dec 29, 2017, Hyderabad, India

2017 "My Story", Women in Statistics Conference, **Plenary Talk**, Oct 19-21, SanDiego.

2017 "Reverse engineering of genomic networks by statistical association scores", Dec2, 2017
      International Conference on Advances in Mathematical Sciences 2017, **Plenary Talk**, VIT,
      Velore, India

2017 "Reverse engineering of genomic networks by statistical association scores", Dec2, 2017
      ICAMS 2017, VIT, Veloore, India

2017 "Reverse engineering of genomic networks by statistical association scores",Pondicherry
       University2017, Pondicherry, India, Dec7, 2017

2017 Joint Statistical Meetings, Baltimore, Maryland, July 29-Aug 3, 2017. "Advances and
      Challenges in Single Call RNA-Seq Analysis".

2017 16 Annual International Conference on Critical Assessment of Massive Data Analysis (COSI
      of ISMB 2017), Prague, Prague, Czech Republic, July 22-23, 2017, "Unraveling Bacterial
      Fingerprints of City Subways From Microbiome 16S Gene Profiles**".**

2017 61st World Statistics Congress, Marrakech, Morocco, 16-21 July 2017. "Network Analysis of
Next-Generation Sequencing Count Data".

2017 Workshop on Statistical Inference for Biomedical Big Data, Gainesville, USA, April 7-8, 2017,
"Statistical association based analysis for genomic data".

2017 International Statistical Institute Regional Statistics Conference, Bali, Indonesia, March 20-
24, 2017. "Statistical association based analysis for genomic data".

2016

10th ICSA International Conference on Global Growth of Modern Statistics, Shanghai, China,
December 19-22, 2016. "Improved Protein Inference from Tandem Mass Spectrometry Data"

Women in Statistics and Data Science Conference, Crowne Plaza Charlotte Executive Park, North
Carolina, Oct 20-22, 2016. "Advancing Omics Data Analysis: It is good to be wanted"

3rd conference of the International Society for Non-Parametric Statistics (ISNPS), Avignon,
France, June 11-16, 2016. "Differential Network Analysis with Multiply Imputed Lipidomic Data".

2016 Frontiers in Applied & Computational Mathematics (FACM), New Jersey   Institute of
Technology, New Jersey, June 3 - 4, 2016. "Differential Network Analysis with Multiply
Imputed Lipidomic Data"

The 2016 IISA International Conference on Statistics Statistical and Data Sciences, OSU Learning Innovation Center, Oregon State University, Corvaliis, August 18-21, 2016. "Differential Network Analysis with Multiply Imputed Lipidomic Data"

The 2016 IISA International Conference on Statistics Statistical and Data Sciences, OSU Learning Innovation Center, Oregon State University, Corvaliis, August 18-21, 2016. "Invited panel on career development of young statisticians".

Joint Statistical Meetings, Chicago, Illinois: "Gender Realities in Cyberspace". July 30- Aug. 4.

The 1st International Statistical Conference in Croatia – ISCCRO'16. "Improved Protein Inference from Tandem Mass spectrometry Data". May 6, 2016, Zagrev, Croetia.

Clinical & Pharmaceutical Solutions through Analysis, March 13-15, 2017, The UF Clinical & Translational Science Institute, Gainesville, FL. "Differential Network Analysis with Multiply Imputed Lipidomic Data"

ENAR 2016 Spring Meeting : Discussant: "Analysis of Next-Generation Sequencing Data: Increasing Accuracy and Novel Applications". March 6-9.

2015

Ninth International Triennial Calcutta Symposium on Probability and Statistics: "Improved Protein Inference from Tandem Mass Spectrometry Data". December 28 - 31, 2015 at the Department of Statistics, Calcutta University, Kolkata, India.

Kolkata International Conference MSAST 2015. "Improved Protein Inference from Tandem Mass Spectrometry Data". Dec 21-23, 2015

Bioinformatics: 14th Annual conference of Critical Assessment of Massive Data Analysis (CAMDA) 2015, "Exploring the Importance of Cancer Pathways by Meta-Analysis of Differential Protein Expression Networks in Three Different Cancers". July 10-11, 2015.

XIV EMR Brazilian School of Regression Models. "Nonparametric regression using partial least squares dimension reduction in multistate models" Campinas, Brazil, March 2-5, 2015.

"Gender Issue" What Is That? Presentation at the session on "Advancing Women in Statistics: investigating and redressing the gender imbalance", 60th ISI World Statistics Congress I, Brazil 2015.

2014

Bioinformatics: Opening Workshop of 2014-15 Program on Beyond Bioinformatics: Statistical and Mathematical Challenges, "An Integrative Exploratory Analysis of –Omics Data from the ICGC Lung Cancer Data". September 8-12, 2014 SAMSI and North Carolina Biotechnology Center, NC.

Bioinformatics: Opening Workshop of 2014-15 Program on Beyond Bioinformatics: Statistical and Mathematical Challenges, Educational Panel: "Role of the Mathematical Sciences in Bioinformatics Education". September 8-12, 2014 SAMSI and North Carolina Biotechnology Center, NC.

Joint Statistical Meeting "Improved protein inference from tandem mass spectrometry data", Boston, Massachusetts • August 2–7, 2014.

ENAR (International Biometric Society) Spring Meeting, "Differential Network Analysis Using

Microarray Gene Expression Data" March 16 – 19, 2014 Baltimore, Maryland

Featured Research Talk "Differential Network Analysis, Spring Opportunities for Women (2013-14 Education and Outreach Program) at *NimBios*, University of Tennessee, Knoxville, TN, April 9, 2014 (co-sponsored by SAMSI, and is ARA-eligible by NISS/SAMSI University Affiliates.

Featured Panel presentation "Support for Your Work: Going for Grants" Spring Opportunities for Women (2013-14 Education and Outreach Program) at *NimBios*, University of Tennessee, Knoxville, TN, April 10, 2014.

Featured Panel presentation "Build Your Professional Network: NSF Math Institutes and Professional Organizations" at *NimBios*, University of Tennessee, Knoxville, TN, April 10, 2014.

Invited Panel "Graduate Education for the Next Generation" First Conference of Celebrating Women in Statistics; Raleigh, NC May 15-17, 2014.

Invited Talk "Transforming Healthcare through Data" First Conference of Celebrating Women in Statistics; Raleigh, NC May 15-17, 2014.

2013

"Nonparametric regression and partial least squares dimension reduction in multistate models" 2013 ICSA International Conference, Hong Kong, China, December 20 - 23, 2013.

22nd International Workshop on Matrices and Statistics, Toronto, Canada, August 12-15, 2013.

Joint Statistical Meeting "Statistical Inference of Protein Identification Using Tandem Mass Spectrometry Data", Montreal, Canada, August 4, 2013.

"Nonparametric regression and partial least squares dimension reduction in multistate models" Statistics and Its Interactions with Other Disciplines (SIOD 2013), Ho Chi Minh City, Vietnam, June 5-7, 2013.

"Similarity in Network Structures for in vivo and in vitro Data from the Japanese Toxicogenomics Project" (CAMDA 2013), Berlin, July 19-20, 2013.

2012

Eighth International Triennial Calcutta Symposium on Probability & Statistics, Kolkata, India, December 27-30, 2012.

"DelPTM: A Statistical Algorithm to Identify Post Translational Modifications from Tandem Mass Spectrometry (MS/MS) Data" (invited but delivered as topic contributed), JSM, Aug 31, 2012.

"Nonparametric regression using partial least squares dimension reduction in multistate models", (Special Session on High-Dimensional Data), LINSTAT – IWMS'2012, Będlewo, Poland, 16–20 July 2012, July 18, 2012.

2011

Breakfast round table for academic women led by **Datta**, **S.**, for 40th Anniversary of Caucus for Women in Statistics, Joint Statistical Meeting, Miami Beach, Florida, July, 31, 2011.

"Nonparametric regression using partial least squares dimension reduction in multistate models" Statistical Concepts and Methods for the Modern World, Colombo, Sri Lanka, December 28-30,

2011.

"Nonparametric regression using partial least squares dimension reduction in multistate models" 4th International Conference of the ERCIM WG on Computing & Statistics, University of London, UK, December 17-19, 2011.

"Predicting Patient Survival from Proteomic Profile using MALDI-TOF Mass Spectrometry Data", Applied Stochastic Models and Data Analysis (ASMDA 2011), Rome, Italy, June 7 - 10, 2011.

2010

"Rank Aggregation and its Use in Bioinformatics Problems", The First International Conference on Theory and Application of Statistics, Dhaka, Bangladesh, 26-28 December 2010.

"Rank Aggregation and its Use in Bioinformatics Problems". LinStat 2010, Tomar, Portugal, July 27, 2010.

"Monoisotopic Peak Detection and Disease Classification for Mass Spectrometry Data". ENAR Spring Meeting, New Orleans, March 22, 2010.

"Improved Automated Monoisotopic Peak Detection Method for Mass Spectrometry Data", UT-ORNL-KBRIN Bioinformatics Summit, Lake Barkley State Resort Park, Cadiz, KY. March 19, 2010.

"Predicting Survival from High Dimensional Data", International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management, February 10, 2010, Beer Sheva, Israel.

2009

"Improved Automated Monoisotopic Peak Detection Method for Mass Spectrometry Data", Seventh International Triennial Calcutta Symposium, Kolkata, Dec. 28, 2009, Kolkata, India.

"Improved Automated Monoisotopic Peak Detection Method for Mass Spectrometry Data", International Conference on Frontiers of Interface between Statistics and Sciences, Hyderabad, India, 31 December 2009.

"A Statistical Framework for Differential Network Analysis (DNA) from Microarray Data Using Partial Least Squares", FACSS, October 18-22, 2009, Louisville, KY.

"Monoisotopic Peak Detection and Detecting Protein-Protein Interaction", First Institute of Mathematical Statistics Asia Pacific Rim Meeting, Seoul, June 28-July 1, 2009.

2008

"Reverse Engineering To Construct Protein-Protein Interaction Network", JSM August 1-7, 2008, Denver, Colorado.

"Construction of Genetic Association Networks", International Indian Statistical Association Conference: Frontiers of Probability and Statistical Science, May 22-26, 2008, University of Connecticut-Storrs.

"Determination of Optimal Clustering Algorithm by Weighted Rank Aggregation: Cross Entropy Algorithm", UT-ORNL-KBRIN Bioinformatics Summit 2008, March 28, 2008, Cadiz, KY.

2007

International Conference on Statistics, Probability and Related Areas by IISA, January 2-5, 2007, Cochin, India.

2006

International Conference on Multivariate Statistical Methods, Dec 28-29, 2006, Kolkata, India.

"Combining Functional Information in Validation of Statistical Clustering", International Multi-Symposiums on Computer and Computational Sciences (IMSCCS|06), June 20-24, 2006, Zhejiang University, Hangzhou, China.

"Clustering Microarray Data", UT-ORNL-KBRIN Bioinformatics Summit 2006, April 21-23, 2006, Cadiz, Kentucky.

2005

"Feature Selection in Mass Spectrometry Data for Cancer Classification", SCMA 2005 / FIM XII, International Conference on Statistics, Combinatorics, Mathematics and Applications: 12th Annual Conference of the Forum for Interdisciplinary Mathematics, December 2-4, 2005, Auburn University, AL, USA.

"Selecting an Appropriate Clustering Algorithm for Analyzing Microarray Data", Joint Annual Meeting of the Interface and the Classification Society of North America, June 8, 2005 - June 12, 2005, Washington University School of Medicine, St. Louis, Missouri.

International Conference on Future of Statistical Theory, Practice and Education, December 29, 2004 - January 1, 2005, Hyderabad, India.

2004

"Empirical Bayes Screening (EBS) of Many P-values with Applications to Microarray Studies", Mathematical Biosciences Institute, Ohio State University. Invited presenter in the workshop titled "Genomics, Proteomics, and Bioinformatics", October 13th 2004.

Eleventh International Conference on Interdisciplinary Mathematical and Statistical Techniques, SCRA 2004, December 27-29, 2004, Lucknow, India.

"Parametric and Nonparametric Empirical Bayes Adjustments to Multiple P-values for the Detection of Differentially Expressed Genes in Microarray Experiments", JSM August 7-12, 2004, Toronto, Canada.

"Empirical Bayes Screening of Many P-values with Applications to Microarray Studies", Microarray Data Analysis Conference arranged by Infocast Inc. Conference arranged, June 28-29, 2004, Rockville, MD, USA.

"Empirical Bayes Screening of Many P-values with Applications to Microarray Studies", International Conference on Statistics in Health Sciences, June 23-25, 2004, Nantes, France.

"Empirical Bayes Analyses of Multiple p-values For the Detection of Differentially Expressed Genes in Microarray Experiments", IISA 2004 Meeting, May 7-9, 2004, Athens, Georgia, USA.

The Second Asia Pacific Bioinformatics Conference 18-22 Jan, 2004, Dunedin, New Zealand (full paper).

2002

"Comparisons of Clustering Algorithms for Grouping Genes Based of Expression Profiles", SCRA 2002-FIM IX: Ninth International Conference of Forum for Interdisciplinary Mathematics

on Statistics Combinatorics and Related Areas, Department of Statistics and Department of Mathematics: University of Allahabad, Allahabad, UP 211 002, India, December 21-23, 2002.

"Statistical Techniques to Analyze Microarray Data: A Partial Overview", Applying Bioinformatics, from Genes to Systems: The University System of Georgia Research Symposium, Georgia State University, Atlanta, GA, October 3-4, 2002.

"Clustering Algorithms for Microarray Data: Overview and Comparative Studies", International Conference on Current Advances and Trends in Nonparametric Statistics, Crete, Greece, July 2002.

"Clustering Algorithms for Microarray Data: Overview and Comparative Studies", ISA International Conference on Statistics, Probability and Related Areas, Dekalb, Illinois, June 2002.

2001

"Use of Partial Least Squares in Microarray Data", International Conference on Statistics, Combinatorics and Related Areas And The Eighth International Conference of the Forum for Interdisciplinary Mathematics, Wollongong, Australia, December 2001.

"Microarray Data and Bioinformatics: A Statistical Future?" IISA-JSM-2000-2001 Conference, New Delhi, India, January 2001.

2000

"Microarray Data and Bioinformatics: A Statistical Future?", Triennial Calcutta Symposium, Calcutta, India, December 2000.

"Estimation of Selection Parameters Using Multigeneration Cytonuclear Data", Symposium on Inference for Stochastic Processes, Athens, Georgia. Estimation of Selection Parameters University of Georgia, Athens, May 2000.

1999

Sixth International Conference of Forum for Interdisciplinary Mathematics on Statistics, Combinatorics and Related Areas, Mobile, Alabama, December 1999.

1998

IISA International Conference on Statistics: IISA98, Hamilton, Canada, October 1998.

1997
Pathways to the Future workshop, Dallas, Texas, June 1997.

AMS - IMS - SIAM Joint Summer Research Conference on Statistics in Molecular Biology, Seattle, Washington, June 1997.

1995

Second IMS new researchers' meeting, Kingston, Canada, July 1995.

SRCOS/ASA summer conference, Melbourne, FL, Student Paper Award, June 1995.

INVITED COLLOQUIA AND SEMINARS

2019

"Advances and Challenges in Single Cell RNA-Seq Analysis", University of Cincinnati, Cincinnati, Ohio, April, 12, 2019.

2018
"Advances and Challenges in Single Cell RNA-Seq Analysis", NIEHS/NIH, Durham, NC 2/22/2018

"Statistical association based analysis for genomic data" Institute of Bioinformatics, University of Georgia, Athens, GA, April 6, 2018

2017
"Reverse engineering of genomic networks by statistical association scores" Department of Biostatistics, University of Miami, October 31, 2017

"Statistical association based analysis for genomic data", Dental School, University of Florida, September 28, 2017

2016

"Improved Protein Inference from Tandem Mass Spectrometry Proteomics Data" The National University of Uzbekistan, Tashkent,Uzbekistan, April 27, 2016

2015

"Improved Protein Inference from Tandem Mass Spectrometry Data", Nov. 12, Department of Statistics, University of Florida

2014

"Improved Protein Inference from Tandem Mass Spectrometry Data", Nov. 23, Department of Biostatistics, University of Florida

2013

"Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies", Jun 24, Institute of Mathematical Sciences, University of Malaya, Malaysia

"Rank Aggregation for Clustering and Classification", Nov. 7, Department of Statistics, Western Michigan University, Kalamazoo, USA

"Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies", Dec. 26, Department of Statistics, Tunghai University, Taichung, Taiwan

2011

"Novel Ensembling Technique for Mining Complex High Dimensional Data", Steklov Mathematical Institute of Academy of Sciences, St. Petersburg, Russia, June 17, 2011.

"Novel Ensembling Technique for Mining Complex High Dimensional Data", School of Public Health, University of Tampere, Tampere, Finland, June 13, 2011.

"Novel Ensembling Technique for Mining Complex High Dimensional Data", School of Public Health and information Sciences, University of Louisville, May 26, 2011.

Presentation of a Chair's Vision, School of Public Health and Information Sciences, May 26 2011.

2010

"High Dimensional Data: A New Paradigm of Biomedical Research". Research Incubation Meeting, SPHIS, University of Louisville, Louisville, Kentucky, 10 March, 2010.

2008

"Bioinformatics in Cancer Research: A Hype or Hope?" JGB Cancer Center, University of Louisville, March 2008.

Department of Quantitative Health Sciences, Cleveland Clinic, January 2008.

2007

Department of Statistics, Bioinformatics Seminar, Texas A&M University, November 2007.

National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, April 2007.

Department of Epidemiology, Michigan State University, East Lansing, March 2007.

2006

Department of Statistics, University of Kentucky, Lexington, October 2006.

2005

ASA, Kentucky Chapter, December, 2005.

"Statistical Techniques to Analyze Microarray Data: A Partial Overview", BiComB Seminar, SPHIS, Emory University, Atlanta, GA.

"Analysis of MALDI-TOF Data", Proteomics Center, University of Louisville, Louisville, November 2005.

2004

Department of Statistics and Applied Probability, National University of Singapore, December 2004.

Department of Biostatistics and Bioinformatics, University of Louisville, December 2004.

Department of Biostatistics, Univ. of Minnesota, April 2004.

2003

Department of Mathematics, Univ. of North Carolina at Charlotte, April 2003.

Department of Mathematics, Georgia Institute of Technology, March 2003.

Department of Mathematics and Statistics, Georgia State University, March 2003.

2001

A one day professional training course on Poisson regression, Centers for Disease Control and Prevention, Atlanta, GA, February 2001.

Department of Biostatistics, University of Alabama, Birmingham, AL, August 2001.

Department of Statistics, University of Georgia, Athens, GA, April 2001.

2000

Bioinformatics group, Jonathan Arnold Lab, Department of Genetics, University of Georgia, Athens, GA, December 2000.

A two day professional training course on Smoothing using S+, Centers for Disease Control and Prevention, Atlanta, GA, November 2000.

1999

Computer Science Unit and Mathematical Science Unit, Indian Statistical Institute, Calcutta, India, June 1999.
Department of Mathematics and Statistics, University of North Carolina, Charlotte, April 1999.
Department of Statistics, Purdue University, April 1999.

1997

Department of Mathematics and Computer Science, Georgia State University, March 1997.

1996

Department of Genetics, University of Georgia, December, 1996.

1995

Population Biology Group, Emory University, November, 1995.

Department of Biostatistics, Emory University, October, 1995.

Department of Mathematics & Statistics, University of North Florida, April 1995.

REFEREED TALKS AT PROFESSIONAL MEETINGS

2011

"Predicting Patient Survival from Proteomic Profile using MALDI-TOF Mass Spectrometry Data in Non-small Cell Lung Cancer Patients", RECOMB Satellite Conference on Computational Proteomics, March 11-13, 2011, University of California, San Diego, La Jolla, California.

2009

"Next Generation Sequencing: Statistical Challenges and Opportunities", CAMDA 2009, October 5-6, 2009, IL, Chicago.

"Improved Monoisotopic Peak Detection in Mass Spectrometry Data", MCP 2009, Tokyo, Japan, March 27, 2009.

2008

"A Statistical Framework for Differential Network Analysis (DNA) from Microarray Data using Partial

Least Squares", Rocky '08, 6th Annual Rocky Mountain Bioinformatics Conference, Snowmass, CO, December 5-8, 2008.

2007

"Fetal Alcohol Syndrome Detected through Proteomics Screening of the Amniotic Fluid in High-Risk (B6J) and Low-Risk (B6N) C57BL/6 Mice", Teratology Society 47th Annual Meeting, Pittsburg, PA, June 2007.

CONTRIBUTED AND TOPIC CONTRIBUTED TALKS AT PROFESSIONAL MEETINGS

2013

(**Topic Contributed**) JSM Statistical Inference of protein identification using tandem mass spectrometry data, Montreal, Canada, August 2013.

ENAR Spring Meeting DELPTM: A Statistical Algorithm to Identify Post-Translational Modifications from Tandem Mass Spectrometry (MS/MS) Data, Orlando, March 10-13, 2013.

2004

ENAR Spring Meeting, Pittsburgh, March 2004.

2003

Digital Biology: The Emerging Paradigm, NIH/BISTI, Bethesda, Maryland, November 2003.

1999

Joint Statistical Meetings, Baltimore, Maryland, August 1999.

ENAR Spring Meeting, Atlanta, Georgia, March 1999.

1998

19th International Biometric Conference: IBC-98, Cape Town, South Africa, December 1998.

1996

18th International Biometric Conference: IBC-96, Amsterdam, The Netherlands, June 1996.

IMS/ENAR meeting, Richmond, Virginia, March 1996.

1995

ASA Winter Conference, Raleigh, North Carolina, January 1995.

REFEREED POSTER PRESENTATION IN PROFESSIONAL AND RESEARCH MEETINGS

2008

"Stochastic Modeling and Statistical Inference of Time Course Metabolic Data". Integrative BioSystems Institute Conference, October 18 -21, 2008, Atlanta, Georgia.

2007

"Finding Cancer Genes Through Meta-analysis of Microarray Experiments: Rank   Aggregation via the Cross Entropy Algorithm". CAMDA 2007, December 2007, Valencia, Spain.

"clValid: An R package for Cluster Validation". July 2007, ISMB 2007, Vienna, Austria.

2005

"Selecting a Clustering Algorithm for Statistical Consistency and Biological Relevance for Gene Expression Data". June 27, 2005, ISMB 2005, Detroit, Michigan.

MEETINGS ATTENDED

2011

Steering committee meeting for the "Celebration of Women in Statistics" in 2013, ASA, Alexandria, VA, Nov.4, 2011.

7th Annual Conference of US HUPO. Proteomics: New Developments and Grand Challenges, March 20 - 23, 2011 - Raleigh, NC

2007

Joint Statistical Meetings, Washington DC, August 2007.

2006

Joint Statistical Meetings, Seattle, Washington, August 2006.

International Biometric Society Conference IBC 2006, Montreal, Canada, July, 2006.

2003

Joint Statistical Meetings, San Francisco, California, August 2003.

Joint Statistical Meetings, Atlanta, Georgia, August 2003.

2001

Workshops for Junior Biostatisticians, 2001 ENAR, Charlotte, March 2001.

2000

Genomics and Medicine Symposium, Emory University, Atlanta, September 2000.

Bioinformatics Conference, Georgia Research Alliance, Atlanta, June 2000 (invited).

Beyond Genome 2000, San Francisco, June 2000.

1998

Joint Statistical Meeting, Dallas, Texas, August 1998.

TEACHING

@University of Florida
1. Advanced Biostatistical Methods II, Spring 2016-2019
2. Frontier's in Biostatistics (Guest Lecture)

@ University of Louisville

Graduate Level Only

1. Mathematics Statistics
2. Biostatistical Consulting
3. (Developed and taught) Statistics for Bioinformatics
4. (Developed and taught) High-throughput Data Analysis
5. (Developed and taught) Statistics for Proteomics
6. (Developed and taught) Multivariate Statistics for Genomics and Proteomics
7. (Developed and taught) Stochastic Modeling and Statistical Inference for Time-Course Data
8. (Developed and taught) Data mining for Genomics and Proteomics
9. Survival Analysis
10. (Developed and Taught) Statistics for Omics
11. (Developed and Taught) Statistical Genetics

@ Georgia State University

Graduate

1. Introductory Biostatistics
2. (Developed and Taught) Statistics for Bioinformatics
3. (Developed and taught) Survival Analysis
4. (Developed and taught) Statistical Computing with S+

Undergraduate

1. Introductory Biostatistics
2. (Developed and taught) Developed an interdisciplinary course on "Scientific Perspectives on World Hunger" Two other faculty from Nutrition and Geology – This course is used as one of the model courses in Models of Accessible Course Design under the GRADE project conducted by Georgia Institute of Technology (GT), IDET Communication Inc. (IDET), and the Southeast Disability and Business Technical Assistance Center (Southeast DBTAC).

@ University of Georgia

Undergraduate

2000   Summer long course for "Analysis of microarray data with R" for NSF/REU program, Summer, 2000.

TEACHING OUTSIDE OF CURRENT INSTITUTION

@ Center for Disease Control and Prevention (CDC), Atlanta, GA

2001   (Developed and taught) Professional training course on Poisson regression
2000   (Developed and taught) A two day professional training course on Smoothing using

S+

POSTDOCTORAL DIRECTION

2007-8 Grzegorz M Boratyn jointly with John Klein and Somnath Datta (Currently at NLM, National Institute of Health).

2013 (for three months) Visiting Scientist Dwijesh Mishra, Indian Agricultural Statistics Research Institute, Centre for Agricultural Bio-Informatics (CAB-in), India.

STUDENT DIRECTION

Major or Co-Major Professor

International Student advising:

2013- Maiju Elisa Kujala Jointly with Jaakko Nevalainen at University of Turku, Finland

PhD (@ University of Florida)

2019-	Runzhe Zhang
2018-	Mingjing Li, Expected Completion, May 2021 (Joint with Yang Yang)
2017-	Dorothy Ellis, Expected Completion, May 2023 (joint with Yang Yang)
2016	Sinjini Sikdar, Completed, May 2017 (Post-Doc at NIEHS/NIH).
2017	Mi Xinlei, Expected completion, May 2018.
2016	Yeonil Kim, Completed, May 2019.
2016	Yichen Chi, Completed, May 2019

MS (@ University of Floria)

2018	Xiao Yi, May, 2017
2019	Runzhe Zhang, May 2019

MPH (@University of Florida)

2016   Li Rina, Completed, May 2017.

PhD (@ University of Louisville)

2015- You Wu, (jointly with Maiying Kong) Completed, May 2017 (Past post-Doc at UC San Francisco followed by Biostatistician's position at Genentech.

2015- Michael Sekula (jointly with Jeremy Gaskins) Expected completion, May 2020.

2013- Maiju Kujala/Pesonen, PhD student (jointly with Jaakko Nevalainen), Turku University, Finland,   Completed Dec. 2016 (Currently a post-Doc at . Aalto University, Helsinki Institute for Information Technology HIIT, Helsinki

2014   Yubing Wan (jointly with Maiying Kong), Completed, August 2014 (Currently Statistician at Precision For Medicine, Frederick, Maryland).

2009   Mourad Atlas, completed, Summer 2009 (Joined FDA, MATHEMATICAL STATISTICIAN).

2009   Vasyl Pihur (jointly with Somnath Datta), completed Summer 2009. (Privacy/Security Engineering Manager at Snapchat, Inc.).

2010   Farida Mostazabi (jointly with Somnath Datta), Graduated, June 2011. (Currently at Medpace Inc., Cincinnati, OH 45227.)

2012   Juliet Ndukum, Completed, June 2012.

2013   Sutirtha Chakraborty (jointly with Somnath Datta), Completed, June 2013. (Currently Manager at Novartis, India)

MS (@ University of Louisville)

2014   Sekula, Michael Nathan, Graduation with MS, May 2015.
2007   Christopher N. Barnes, Feature Selection and Classification in High-throughput Data Analysis, graduated summer 2007 (in industry after PhD).
2008   Becky Patterson (jointly with P. Yoo), Testing the Effects of Predictors Using Data Generated by Non-identity Link Functions of the Single-Index Model: A Monte Carlo Approach, graduated spring 2008 (Employed in Graduate School UofL).
2010   Xiaohong Li, Completed, Spring, 2010 (Enrolled in the PhD program in Biostatistics at UofL).
2011   Jasmit Shah, completed, Summer, 2011, (Enrolled in the PhD program in Biostatistics at UofL).

MS (@ Georgia State University)

2001   Sylvie Bougi, A Comparative Power Study of Statistical Tests of Neutrality of DNA Markers Using Multigeneration Cytonuclear Data; graduated Spring 2001.
2002   Israel Hora, Time Series Analysis of Georgia Employment Data and Future Prediction of Employment Status, graduated, Fall, 2002.
2002   Ying, Yang (Biology), non-thesis report, graduated Summer, 2002.
2004   Jennifer Le-Rademacher, Partial Least Squares in Censored Survival Regression, graduated Summer 2004.
2004   Xiao Hong Zhu, graduated Summer 2004.
2004   Usha Ramakrishnan, graduated Fall 2004.
2005   Mourad Atlas, graduated Spring 2005.
2005   Lace Depadilla, graduated Spring 2005.
2005   Baofu Ma, graduated Summer 2005.
2005   Saurav Karmaker, graduated Summer 2005.

Committee Member at University of Louisville

2014   Sandipan Dutta, PHD, Expected Graduation May, 2016
2014   Hyoyoung Choo-Wosoba, Expected Graduation May, 2016
2014   Joseph Bible, PHD, Expected Graduation May, 2015
2011   Stephanie Tabb, PHD, Graduated Fall 2011.
2011   Hyejeong Jang, MS, Graduated May 2011.
2008   Ling Lang PHD, Graduated August 2008.
2007   Vasyl Pihur MS. Graduated August 2007.

Visiting Research Trainee (from abroad)

2013 Dwijesh Chandra Mishra, PhD, Centre for Agricultural Bioinformatics (CABin), Indian Agricultural Statistics Research Institute, New Delhi and a faculty (Bioinformatics) in Post Graduate School at Indian Agricultural Research Institute (Deemed University), Pusa, New Delhi, India.July-September, 2013

2013 Maiju Kujala, PhD student, Turku University, Finland, August-November, 2013

Faculty Mentoring (Official at University of Florida)

Oral Biology, School of Dentistry, University of Florida (Jorge Frias-Lopez, Associate Professor)