

## Debashis Ghosh

2564 Forest Street  
Denver, CO 80207

Colorado School of Public Health  
Department of Biostatistics and Informatics  
Bldg. 500, 3rd Fl, Room W 4134  
13001 East 17th Place  
Aurora, CO 80045  
Phone: (303) 724-4365  
E-mail: debashis.ghosh@cuanschutz.edu  
Web: <http://github.com/GhoshLab/>

### Education:

2000	University of Washington	Ph.D., Biostatistics
1997	University of Washington	M.S., Biostatistics
1995	Rice University, <i>summa cum laude</i>	B.A., Mathematics/Statistics B.A., Economics/French Studies

### Academic Positions:

2017-2019	Analytics Core Co-Director University of Colorado Data Science to Patient Value Initiative
2016	Interim Director, Biostatistics and Bioinformatics Shared Resource University of Colorado Comprehensive Cancer Center
2014-2019	Associate Director, Colorado Center for Biomedical Informatics and Personalized Medicine University of Colorado Anschutz Medical Campus, Aurora, CO
2014-present	Chair, Department of Biostatistics and Informatics Colorado School of Public Health, Aurora, CO
2014-present	Professor, Department of Biostatistics and Informatics Colorado School of Public Health, Aurora, CO
2011-2014	Investigator, The Methodology Center, Penn State University University Park, PA
2010-2014	Professor, Departments of Statistics and Public Health Sciences, Penn State University, University Park, PA
2009-2011	Adjunct Associate Professor, Department of Biostatistics, University of Michigan, Ann Arbor
2008-2014	Member, Center for Comparative Genomics and Bioinformatics Penn State University
2008-2014	Affiliate Faculty, Penn State Cancer Institute Penn State College of Medicine, Hershey
2007-2010	Associate Professor, Department of Public Health Sciences, Penn State College of Medicine, Hershey
2007-2010	Associate Professor, Department of Statistics, Penn State University, University Park

2007-2014 Affiliate Faculty, Integrated Biosciences (IBIOS) Program  
Penn State University, University Park

2005-2007 Associate Professor, Department of Biostatistics,  
University of Michigan, Ann Arbor

2003-2007 Assistant Member, UM Comprehensive Cancer Center

2002-2007 Affiliate Faculty, Bioinformatics Program  
University of Michigan, Ann Arbor

2001-2007 Affiliate Faculty, Center for Statistical Genetics  
University of Michigan, Ann Arbor

2001-2005 Assistant Professor, Department of Biostatistics  
University of Michigan, Ann Arbor

2000 Postdoctoral Scholar, Institute for Pure and Applied Mathematics,  
University of California, Los Angeles

1995–2000 Research and Teaching Assistant, Department of Biostatistics,  
University of Washington

1996 Statistical Intern, Division of Clinical Statistics  
Abbott Laboratories, Waukegan, IL

#### Awards and Honors:

2022-2024 Editor, *Sankhya B*

2021 University of Colorado Anschutz Medical Campus,  
Outstanding Research Collaboration Award

2020 Sigma Xi (Scientific Honor Society)

2018 Delta Omega (Public Health Honor Society)  
Colorado School of Public Health

2018 Outstanding Young Statistical Scientist Award, Applications Track  
International Indian Statistical Association

2018 – 2020 Co-Editor, *Biometrics*

2015 Myrto Lefkopolou Distinguished Lecturer, Department of Biostatistics  
Harvard T.W. Chan School of Public Health

2015 University of Colorado Clinical and Translational Sciences Leadership  
in Innovative Team Science Program Participant

2014 University of Washington Department of Biostatistics Distinguished Alumni Speaker

2014 – present Grohne-Stepp Endowed Chair in Cancer Research  
University of Colorado Cancer Center

2013 – 2015 Chair, Biostatistical Methods and Research Design Study Section,  
National Institutes of Health

2013 Mortimer Spiegelman Award, American Public Health Association

2012 Fellow, American Statistical Association

2012 Featured Cover Article, *Genomics*

2011 Faculty Scholar, Methodology Center, Pennsylvania State University

2010–2012 College of CSR Reviewers, National Institutes of Health

2008 Featured *Biomed Central* Editorial Board Member

2008 *Biometrics* Best Paper Award

2006 UM CCMB Pilot Grant Award

2001-2003 UM Bioinformatics Pilot Grant Award

2002 Pacific Symposium on Biocomputing Travel Award

2001	UM Prostate Cancer SPORE Seed Grant
2001	UM Cancer Center MUNN Idea Grant
1999	ENAR Student Travel Award
1999	Society of Clinical Trials Student Scholarship
1998–2000	NIH Predoctoral Cardiovascular Training Grant, Department of Biostatistics, University of Washington
1997	Donovan J. Thompson Outstanding Student Award for outstanding academic performance, Department of Biostatistics, University of Washington
1995–1997	National Science Foundation Graduate Research Fellowship
1995–1997	ARCS (Achievement Rewards for College Scientists) Fellowship Department of Biostatistics, University of Washington
1995	Phi Beta Kappa, Rice University
1994	Phi Sigma Phi, Rice University

### Books:

Tseng, G., Ghosh, D. and Zhou, X. J. (2015), Editors. *Integrating -Omics Data*. Cambridge: Cambridge University Press.

### Publications which have received peer review:

† denotes a master's student advised by Ghosh, \* denotes a first-authored publication by Ph.D. student or postdoc advised/co-advised by Ghosh.

1. Ghosh, D. and Godbole, A. (1997). Palindromes in random letter generation: Poisson approximations, rates of growth, and Erdős-Rényi laws. In *Proceedings of the Athens Conference on Applied Probability*, ed. C.C. Heyde *et al.*, Springer-Verlag Lecture Notes in Statistics, vol. 114, pp. 99–115.
2. Ghosh, D. and Lin, D.Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554–562.
3. Ghosh, D. (2000). Methods for the analysis of multiple events and death. *Controlled Clinical Trials* **21**, 115–126.
4. Ghosh, D, Deisher, T.A. and Ellsworth, J.E. (2000). Methods for the analysis of repeated measures. *Journal of Pharmacological and Toxicological Methods* **42**, 157–162.
5. Ghosh, D. (2001). Efficiency considerations in the additive hazards model with current status data. *Statistica Neerlandica* **55**, 367 – 376.
6. Sreekumar, A., Nyati, M., Barrette, T. R., Ghosh, D., Lawrence, T. and Chinnaiyan, A. M. (2001). Profiling cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Research* **61**, 7585 – 7593.
7. Dhanasekaran, S., Barrette, T., Ghosh, D., Shah, R., Kurachi, K., Pienta, K., Rubin, M. A. and Chinnaiyan, A. M. (2001). Molecular profiling of prostate cancer: delineation of candidate biomarkers and regulatory genes. *Nature* **412**, 422 – 426.

8. Ghosh, D. (2002). Singular value decomposition regression models for the classification of tumors from microarray experiments. In *Proceedings of the 2002 Pacific Symposium on Biocomputing*, Eds. Altman, R. B. et al. pp. 18 – 29.
9. Ghosh, D. and Chinnaiyan, A. M. (2002). Mixture modelling of gene expression data from microarray experiments. *Bioinformatics* **18**, 275–286.
10. Rubin, M. A., Zhou, M., Dhanasekaran, S. M., Varambally, S., Barrette, T. R., Sanda, M. G., Pienta, K. J., Ghosh, D. and Chinnaiyan, A. M. (2002).  $\alpha$ -methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. *Journal of the American Medical Association* **287**, 1662–1670.
11. Goldstein, D., Ghosh, D. and Conlon, E. (2002). Statistical issues in the clustering of gene expression data. *Statistica Sinica* **12**, 219–241.
12. Ghosh, D. (2002). Resampling methods for variance estimation of singular value decomposition analyses from microarray experiments. *Functional and Integrative Genomics* **2**, 92 – 97.
13. Ghosh, D. and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663 – 688.
14. Rhodes, D., Barrette, T. T., Rubin, M. A., Ghosh, D. and Chinnaiyan, A. M. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Research* **62**, 4427 – 4433.
15. Varambally, S., Dhanasekaran, S. M., Zhou, M., Barrette, T. R., Kumar-Sinha, C., Sanda, M. G., Ghosh, D., Pienta, K. J., Sewalt, R. G. A. B., Otte, A. P., Rubin, M. A. and Chinnaiyan, A. M. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624 – 629.
16. Wu, J., Haan, M., Liang, J., Ghosh, D., Gonzalez, H., Jagust, W. and Herman W. (2003). Impact of antidiabetic medications on physical and cognitive functioning of older Mexican Americans with diabetes mellitus. *Annals of Epidemiology* **13**, 369 – 376.
17. Wu, J. H., Haan, M. N., Liang, J., Ghosh, D., Gonzalez, H. M. and Herman W. H. (2003). Diabetes as a predictor of change in functional status among older Mexican Americans: a population-based cohort study. *Diabetes Care* **26**, 314 – 319.
18. Wu, J. H., Haan, M. N., Liang, J., Ghosh, D., Gonzalez, H. M. and Herman W. H. (2003). Impact of diabetes on cognitive function among older Latinos: a population-based cohort study. *Journal of Clinical Epidemiology* **56**, 686 – 693.
19. Ghosh, D. (2003). Goodness of fit methods for the additive risk model in tumorigenicity experiments. *Biometrics* **59**, 721 – 726.
20. Ghosh, D., Barrette, T., Rhodes, D. and Chinnaiyan, A. M. (2003). Statistical issues and procedures for meta-analysis of microarray data: a case study in prostate cancer. *Functional and Integrative Genomics* **3**, 180 – 188.

21. †Smolkin, M. and Ghosh, D. (2003). Cluster stability scores for cancer subtypes in microarray experiments. *BMC Bioinformatics* **4**, 36 – 42.
22. Kleer, C. G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S. A., Ghosh, D., Sewalt, R. G., Otte, A. P., Hayes, D. F., Sabel, M. S., Livant, D., Weiss, S. J., Rubin, M. A. and Chinnaiyan, A. M. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proceedings of the National Academy of Sciences USA* **100**, 11606 – 11611.
23. Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events in the presence of dependent censoring. *Biometrics* **59**, 877 – 885.
24. Ghosh, D. (2003). Penalized discriminant methods for the classification of tumors from microarray experiments. *Biometrics* **59**, 992 – 1000.
25. Sreekumar, A., Laxman, B., Rhodes, D., Bhagavathula, S., Giacherio, D., Ghosh, D., Sanda, M. G., Rubin, M. and Chinnaiyan, A. M. (2004). Humoral immune response to alpha-methylacetyl-CoA racemase and prostate cancer. *Journal of the National Cancer Institute* **96**, 834 - 843.
26. Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1 – 6.
27. Yu, J., Mears, A. J., Yoshida, S., Farjo, R., Carter, T. A., Ghosh, D., Hero, A., Barlow, C., Swaroop, A. (2004). From disease genes to cellular pathways: A progress report. In “Retinal dystrophies: functional genomics to gene therapy.” Wiley, Chichester (Novartis Foundation Symposium 255), pp. 147-164.
28. Sen, S., Burmeister, M. and Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *American Journal of Medical Genetics B Neuropsychiatric Genetics* **127**, 85 – 89.
29. Zhang, L., Srinivasan, U., Marrs, C. F., Ghosh, D., Gilsdorf, J. R. and Foxman, B. (2004). Library on a slide for bacterial comparative genomics. *BMC Microbiology* **4**, 12.
30. Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A. M. (2004). Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proceedings of the National Academy of Sciences USA* **101**, 9309 – 9314.
31. \*Shen, R., Ghosh, D. and Chinnaiyan, A. M. (2004). Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics* **5**, 94. (**Note:** This paper earned Ronglai Shen an ENAR student travel award in 2003).
32. Ghosh, D. (2004). Mixture models for assessing differential expression in complex tissue using microarray data. *Bioinformatics* **20**, 1663 – 1669.
33. Ghosh, D. (2004). Accelerated rates regression models for recurrent failure time data. *Lifetime*

*Data Analysis* **10**, 247 – 261.

34. Ghosh, D. (2004). Identification of shared components and sparse networks in gene expression data. *Journal of Virtual Learning Signal Integration and Signal Processing* **38**, 277 – 286.
35. Yu, J., He, S., Friedman, J. S., Ghosh, D., Mears, A. J., Hicks, D. and Swaroop, A. (2004). Altered expression of genes of the Bmp/Smad and Wnt/Calcium signaling pathways in the cone-only Nrl-knockout mouse retina, revealed by gene profiling using custom cDNA microarrays. *Journal of Biological Chemistry* **279**, 42211 - 42220.
36. Shah, R.B., Mehra, R., Chinnaiyan, A.M., Shen, R., Ghosh, D., Zhou, M., MacVicar, G.R., Varambally, S., Harwood, J., Bismar, T.A., Kim, R., Rubin, M.A. and Pienta, K.J. (2004). Androgen independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. *Cancer Research* **64**, 9209 - 9216.
37. Ghosh, D. and Chinnaiyan, A. M. (2005). Covariate adjustment in the analysis of microarray data from clinical studies. *Functional and Integrative Genomics* **5**, 18 – 27.
38. Ghosh, D. (2005). Nonparametric methods for identifying multiple replications of origins in genomewide data. *Functional and Integrative Genomics* **5**, 28 – 31.
39. Shedden, K., Chen, W., Kuick, R., Ghosh, D., Macdonald, J., Cho, K., Giordano, T. J., Gruber, S. B., Fearon, E. R., Taylor, J. M. G.G. and Hanash, S. (2005). A comparison of seven methods for normalization with Affymetrix expression scores based on false discovery rates in disease profiling data. *BMC Bioinformatics* **6**, 26.
40. Mallick, B., Ghosh, D. and Ghosh, M. (2005). Bayesian kernel-based classification of microarray data. *Journal of the Royal Statistical Society Series B* **2**, 219 – 234. Republished in *Machine Learning in Bioinformatics*, ed. S. Mitra *et al.*. (2008). New York: Chapman and Hall.
41. Abecasis, G. R., Ghosh, D. and Nichols, T.E. (2005). Linkage disequilibrium: ancient history drives the new genetics. *Human Heredity* **59**, 118 – 124.
42. Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Barrette, T. R., Ghosh, D. and Chinnaiyan, A. M. (2005). Mining for regulatory programs in the cancer transcriptome. *Nature Genetics* **37**, 579 – 583.
43. Witkiewicz, A., Varambally, S., Shen, R., Mehra, R., Sabel, M., Ghosh, D., Chinnaiyan, A. M., Rubin, M. A. and Kleer, C. G. (2005). Alpha-Methylacyl-CoA Racemase (AMACR) protein expression is associated with the degree of differentiation in breast cancer using quantitative image analysis. *Cancer Epidemiology, Biomarkers and Prevention* **14**, 1418 – 1423.
44. Rubin, M. A., Bismar, T. A., Andren, O., Mucci, L., Kim, R., Shen, R., Ghosh, D., Wei, J. T., Chinnaiyan, A. M., Adami, H. O., Kantoff, P. W., and Johansson, J. E. (2005). Decreased Alpha-Methylacyl CoA racemase expression in localized prostate cancer is associated with an increased rate of biochemical recurrence and cancer specific death. *Cancer Epidemiology, Biomarkers and Prevention* **14**, 1424 – 1432.

45. Levin, A. M., Ghosh, D., Cho, K. R. and Kardia, S. L. R. (2005). A model-based scan statistic for identifying extreme chromosomal regions of gene expression in human tumors. *Bioinformatics* **21**, 2867 – 2874.
46. Sen, S., Burmeister, M., and Ghosh, D. (2005). 5-HTTLPR and anxiety-related personality traits meta-analysis revisited: response to Munafò and colleagues. *Molecular Psychiatry* **10**, 893 – 895.
47. Ghosh, D. and Chinnaiyan, A. M. (2005). Classification and selection of biomarkers in genomic data using LASSO. *Journal of Biomedicine and Biotechnology* **2**, 147 – 154.
48. Rhodes, D. R., Tomlins, S. A., Varambally, S., Mahavisno, V., Barrette, T., Kalyana-Sundaram, S., Ghosh, D., Pandey, A. and Chinnaiyan, A. M. (2005). Probabilistic model of the human protein-protein interaction network. *Nature Biotechnology* **23**, 951 – 959.
49. Wang, X, Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., Montie, J., Pienta, K. J., Sanda, M. G., Kantoff, P. W., Rubin, M. A., Wei, J.T., Ghosh, D., and Chinnaiyan, A. M. (2005). Autoantibody signatures in prostate cancer. *New England Journal of Medicine* **353**, 1224 – 1235.
50. Mehra, R., Varambally, S., Shen, R., Hayes, D. F., Sabel, M. S., Ghosh, D., Chinnaiyan, A. M. and Kleer, C. G. (2005). Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Research* **65**, 11259 – 11264.
51. Varambally, S., Yu, J., Laxman, B., Rhodes, D. R., Mehra, R., Shah, R., Chandran, U., Monzon, F. A., Becich, M. J., Wei, J. T., Pienta, K. G., Ghosh, D., Rubin, M. A. and Chinnaiyan, A. M. (2005). Integrative molecular analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393 – 406.
52. McEwen, L. M., Kim, C., Haan, M., Ghosh, D., Lantz, P., Mangione, C., Safford, M., Marrero, D., Thompson, T. and Herman, W. (2006). Diabetes reporting as a cause of death: results from Translating Research Into Action for Diabetes (TRIAD). *Diabetes Care* **29**, 247 – 253.
53. Ghosh, D., Chen, W. and Raghunathan, T. E. (2006). The false discovery rate: a variable selection perspective. *Journal of Statistical Planning and Inference* **136**, 2668 – 2684.
54. Ghosh, D. (2006). Modelling tumor biology-progression relationships in screening trials. *Statistics in Medicine* **25**, 1872 – 1884.
55. Ghosh, D. (2006). Semiparametric inferences for the association parameter with semi-competing risks data. *Statistics in Medicine* **25**, 2059 – 2070.
56. Ghosh, D. (2006). Random projection-based clustering for finding cancer subtypes in cancer microarray data. *Neurocomputing* **69**, 2258 – 2267.
57. Elder, J. T., Ghosh, D. and Shah, R. B. (2006). Epidermal growth factor receptor (ErbB1) expression in prostate cancer progression: correlation with androgen independence. *Prostate* **66**, 1437 – 1444.

58. Tsai, H. J., Tsai, A. C., Nriagu, J., Ghosh, D., Gong, M. and Sandretto, A. (2006). Risk factors for respiratory symptoms and asthma in the residential environment of 5th grade schoolchildren in Taipei, Taiwan. *Journal of Asthma* **5**, 355 – 361.
59. \*Shen, R., Ghosh, D., Chinnaiyan, A. M. and Meng, Z. (2006). Eigengene-based linear discriminant analysis for gene expression data. *Bioinformatics* **22**, 2635 – 2642.
60. Chang, S., Ghosh, D., Linderman, J. L. and Kirschner, D. E. (2006). Length-based prediction of peptide-MHC class II binding affinity. *Bioinformatics* **22**, 2761 – 2767.
61. Macdonald, J. W. and Ghosh, D. (2006). COPA - cancer outlier profile analysis. *Bioinformatics* **22**, 2950 – 2951.
62. Ghosh, D. (2006). Semiparametric global cross-ratio models for bivariate censored data. *Scandinavian Journal of Statistics* **33**, 609 – 619.
63. Banerjee, M., Biswas, P., and Ghosh, D. (2006). Semiparametric binary regression under monotonicity constraints. *Scandinavian Journal of Statistics* **33**, 673 – 697.
64. Ghosh, D. (2006). Shrunken p-values for assessing differential expression, with applications to genomic data analysis. *Biometrics* **62**, 1099 – 1106.
65. Xie, J., Juliao, P., Gilsdorf, J. R., Ghosh, D., Patel, M., McCrea, K. W. and Marrs, C. F. (2006). Identification of more prevalent virulence factors in nontypeable Haemophilus influenzae otitis media strains than in throat strains. *Journal of Clinical Microbiology* **44**, 4316 – 4325.
66. Srinivasan, U., Zhang, L., France, A. M., Ghosh, D., Shalaby, W., Xie, J., Marrs, C. F. and Foxman, B. (2007). Probe hybridization array typing (PHAT): a binary typing method for E. coli. *Journal of Clinical Microbiology* **45**, 206 – 214.
67. Ghosh, D. and Chinnaiyan, A. M. (2007). Empirical Bayes identification of tumor progressor genes from microarray data. *Biometrical Journal* **49**, 68 – 77.
68. Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B. B., Barrette, T. R., Anstet, M. J., Kincaid-Beal, C., Kulkarni, P., Varambally, S., Ghosh, D. and Chinnaiyan, A. M. (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166 – 180.
69. Slotnick, M. J., Meliker, J. R., Avruskin, G. A., Ghosh, D. and Nriagu, J. O. (2007). Toenails as a biomarker of inorganic arsenic intake from drinking water and foods. *Journal of Toxicology and Environmental Health* **70**, 148 – 158.
70. Ghosh, D. (2007). Incorporating monotonicity into the evaluation of a biomarker. *Biostatistics* **8**, 402 – 413.
71. Chen, G., Wang, X., Yu, J., Varambally, S., Yu, J., Thomas, D. G., Wang, Z., Fielhauer, J., Ghosh, D., Giordano, T. J., Giacherio, D., Chang, A. C., Orringer, M. B., Bigsbee, W., Beer, D. G. and Chinnaiyan, A. M. (2007). Diagnostic autoantibodies for lung adenocarcinoma includes



ubiquilin 1. *Cancer Research* **67**, 3461 – 3467.

72. Mehra, R., Tomlins, S. A., Shen, R., Nadeem, O., Wang, L., Wei, J. T., Pienta, K. J., Ghosh, D., Rubin, M. A., Chinnaiyan, A. M. and Shah, R. B. (2007). Comprehensive assessment of TM-PRSS2 and ETS family gene aberrations in clinically localized prostate cancer. *Modern Pathology* **20**, 538 – 544.

73. \*Poisson, L. M. and Ghosh, D. (2007). Statistical issues and analyses of *in vitro* and *in vivo* genomic data in order to identify clinically relevant profiles, *Cancer Informatics* **1**, 231 – 243.

74. Yu, J., Almal, A. A., Yu, J., Dhanasekaran, S. M., Ghosh, D., Worzel, W. P., and Chinnaiyan, A. M. (2007). Molecular classification of cancer and feature selection using genetic programming. *Neoplasia* **9**, 292 – 303.

75. Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B. B., Barrette, T. R., Anstet, M. J., Kincead-Beal, C., Kulkarni, P., Ghosh, D., Varambally, S., and Chinnaiyan, A. M. (2007). The Oncomine concepts map links tumors, pathways, mechanisms and drugs. *Neoplasia* **9**, 443 – 454.

76. McEwen, L. N., Kim, C., Karter, A. J., Haan, M., Ghosh, D., Lantz, P. M., Mangione, C. M., Thompson, T. J. and Herman, W. H. (2007). Risk factors for mortality among patients with diabetes: the translating research into action for diabetes (TRIAD) study. *Diabetes Care* **30**, 1736 – 1741.

77. Davila, M., Jhala, D., Ghosh, D., Grizzle, W. E. and Chakrabarti, R. (2007). Expression of Lim kinase 1 is associated with reversible G1/S arrest, chromosomal instability and prostate cancer. *Molecular Cancer* **6**, 40.

78. Tsai, H. J., Tsai, A. C., Nriagu, J., Ghosh, D., Gong, M., and Sandretto, A. (2007). Associations of BMI, TV-watching time, and physical activity on respiratory symptoms and asthma in 5th grade schoolchildren in Taipei, Taiwan. *Journal of Asthma* **44**, 397 – 401.

79. \*Liu, D., Lin, X. and Ghosh, D. (2007). Semiparametric regression of multi-dimensional genetic pathway data: least squares kernel machines and linear mixed models. *Biometrics* **63**, 1079 – 1088. (**Note:** This paper earned Dawei Liu an ENAR student travel award in 2003 and the *Biometrics* best paper award in 2007).

80. Kim, J. H., Dhanasekaran, S. M., Mehra, R., Tomlins, S. A., Gu, W. J., Yu, J., Kumar-Sinha, C., Cao, X., Dash, A., Wang, L., Ghosh, D., Shedden, K., Montie, J. E., Rubin, M. A., Pienta, K. J., Shah, R. B., and Chinnaiyan, A. M. (2007). Integrative analysis of genomic aberrations associated with prostate cancer progression. *Cancer Research* **67**, 8229 – 39.

81. \*Choi, H., \*Shen, R., Chinnaiyan, A. M. and Ghosh, D. (2007). A latent variable approach for meta-analysis of gene expression data from multiple microarray experiments. *BMC Bioinformatics* **8**, 364.

82. Chakraborty, S., Mallick, B., Ghosh, D., Ghosh, M. and Dougherty, E. (2007). Gene expression-based glioma classification using hierarchical Bayesian vector machines. *Sankhya* **69**, 514 – 547.

83. Yu, J., Yu, J., Cao, Q., Mehra, R., Laxman, B., Creighton, C. J., Tomlins, S. Dhanasekaran, S. M., Zhou, W., Chen, G., Shah, R. B., Ghosh, D., Varambally, S. and Chinnaiyan, A. M. (2007). Integrative genomics analysis reveals silencing of  $\beta$ -adrenergic signaling by polycomb in cancer. *Cancer Cell* **12**, 419 – 431.
84. Yu, J., Yu, J., Rhodes, D. R., Tomlins, S. A., Cao, X., Chen, G., Mehra, R., Dhanasekaran, S. M., Wang, X., Ghosh, D., Shah, R. B., Varambally, S., Pienta, K. J., and Chinnaiyan, A. M. (2007). A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Research* **67**, 10657 – 10663.
85. Taylor, B. S., Pal, M., Yu, J., Laxman, B., Sundaram, S. K., Zhao, R., Menon, A., Wei, J. T., Nesvizhskii, A. I., Ghosh, D., Omenn, G. S., Lubman, D. M., Chinnaiyan, A. M. and Sreekumar, A. (2008). Humoral response profiling reveals pathways to prostate cancer progression. *Molecular and Cellular Proteomics* **7**, 600 – 611.
86. \*Choi, H. W., Ghosh, D. and Neshvizhskii, A. (2008). Statistical validation of peptide identifications in large-scale proteomics using target-decoy database search strategy and flexible mixture modeling. *Journal of Proteome Research* **7**, 286 – 292.
87. Yu, J., Yu, J., Cordero, K. E., Johnson, M. D., Ghosh, D., Chinnaiyan, A. M., Rae, J. M., and Lippman, M. E.. (2008). A transcriptional fingerprint of estrogen in human breast cancer predicts patient survival. *Neoplasia* **10**, 79-88.
88. \*Gu, W. J., Choi, H. W. and Ghosh, D. (2008). Global associations between copy number and transcript mRNA expression: an empirical study. *Cancer Informatics* **4**, 17 – 23.
89. Kwee, L. C., Liu, D., Lin, X., Ghosh, D. and Epstein, M. P. (2008). A powerful and flexible multilocus association test for quantitative traits. *American Journal of Human Genetics* **82**, 386 – 397.
90. Ghosh, D. (2008). Proportional hazards regression for cancer studies. *Biometrics* **64**, 141 – 148.
91. Ghosh, D. (2008). Semiparametric inference for surrogate endpoints with bivariate censored data. *Biometrics* **64**, 149 – 156.
92. \*Shen, R., Ghosh, D. and Taylor, J. M. G.(2008). Analysis of tissue microarray data using measurement error models. *Statistics in Medicine* **27**, 1944 – 1959.
93. \*Chen, W., Ghosh, D., Raghunathan, T. E. and Sargent, D. (2008). A loss function-based framework for simultaneous inference involving interactions. *Statistics in Medicine* **27**, 2004 – 2021.
94. \*Yuan, Z. and Ghosh, D. (2008). Combining multiple biomarker models in logistic regression. *Biometrics* **64**, 431 – 439.
95. Ghosh, D. (2008). On the Plackett distribution with bivariate censored data. *The International Journal of Biostatistics* Vol. 4 : Issue 1, Article 7.

96. \*Liu, D., Ghosh, D. and Lin, X. (2008). Assessing the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via mixed models. *BMC Bioinformatics* **9**, 292.
97. \*Shen, R., Chinnaiyan, A. M. and Ghosh, D. (2008). Pathway analysis reveals functional convergence of gene expression profiles in breast cancer. *BMC Medical Genomics* **1**, 28.
98. Sarkar, S., Zhou, T., and Ghosh, D. (2008). A general decision-theoretic approach to multiple testing procedures for false discovery and false nondiscovery rates. *Statistica Sinica* **18**, 925 – 946.
99. Varambally, S., Laxman, B., Mehra, R., Cao, Q., Dhanasekaran, S. M., Tomlins, S. A., Granger, J., Vellaichamy, A., Sreekumar, A., Yu, J., Gu, W., Shen, R., Ghosh, D., Wright, L. M., Kladney, R. D., Kuefer, R., Rubin, M. A., Fimmel, C. J., and Chinnaiyan, A. M. (2008). Golgi protein GOLM1 is a tissue and urine biomarker of prostate cancer. *Neoplasia* **10**, 1285 – 1294.
100. Colon Lopez, V., Haan, M. N., Aiello, A. E. and Ghosh, D. (2008). Fasting total homocysteine (tHcy) concentration and mortality in older Mexican Americans. *Journal of Nutritional Health and Aging* **12**, 685-689.
101. Ghosh, D., Banerjee, M. and Biswas, P. (2008). Inference for constrained estimation of tumor size distributions. *Biometrics* **64**, 1009 – 1017.
102. \*Shen, R., Taylor, J. M. G. and Ghosh, D. (2008). Reconstructing tumor-wise protein expression in tissue microarray studies using a Bayesian cell mixture model. *Bioinformatics* **24**, 2880 – 2886. (**Note:** This paper won Ronglai Shen a Biometrics Section travel award to JSM 2008).
103. Ghosh, D. and \*Poisson, L. M. (2009). Omics data and levels of evidence for biomarker discovery. *Genomics* **93**, 13 – 16.
104. Ghosh, D. and Chinnaiyan, A. M. (2009). Genomic outlier profile analysis: mixture models, null hypotheses and nonparametric estimation. *Biostatistics* **10**, 60 – 69.
105. Colón-López, V., Haan, M. N., Aiello, A. E. and Ghosh, D. (2009). The effect of age at migration on cardiovascular mortality among elderly Mexican immigrants. *Annals of Epidemiology* **19**, 8 – 14.
106. Ghosh, D. (2009). Empirical Bayes methods for estimation and confidence intervals in high-dimensional problems. *Statistica Sinica* **19**, 125 – 143.
107. Sreekumar, A., \*Poisson, L. M., Rajendiran, T. M., Khan, A. P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R. J., Yong, L., Nyati, M. K., Ahsan, A., Kalyana-Sundaram, S., Han, B., Cao, X., Byun, J., Omenn, G. S., Ghosh, D., Pennathur, S., Alexander, D. C., Berger, A., Shuster, J. R., Wei, J. T., Varambally, S., Beecher, C., and Chinnaiyan, A. M. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910 – 914. (**Note:** A. Sreekumar and L. M. Poisson are co-first authors on this publication.)
108. \*Wu, J., Patwa, T., Lubman, D. M. and Ghosh, D. (2009). Multiplicity-adjusted tests for identifying differential expressed clusters based on spatially correlated proteomic data. *Computational*

109. McEwen, L. N., Kim, C., Haan, M. N., Ghosh, D., Lantz, P. M., Thompson, T. J., and Herman, W. H. (2009). Are health-related quality-of-life and self-rated health associated with mortality? Insights from translating research into action for diabetes (TRIAD). *Primary Care Diabetes* **3**, 37 – 42.
110. Davis, S. W., Potoka, M. A., Brinkmeier, M. L., Carninci, P., Lyons, R. H., MacDonald, J., Fleming, M. T., Mortensen, A. H., Egashira, N., Ghosh, D., Steel, K., Osamura, R. Y., Hayashizaki, Y., Camper, S. A. (2009). Genetics, gene expression and bioinformatics of the pituitary gland. *Hormone Research* **71 Suppl 2**, 101 – 115.
111. Ghosh, D. and Yuan, Z. (2009). An improved model averaging scheme for logistic regression. *Journal of Multivariate Analysis* **100**, 1670 – 1681.
112. Ghosh, D. (2009). On assessing surrogacy in a single-trial setting using a semi-competing risks paradigm. *Biometrics* **65**, 521 – 529.
113. \*Choi, H. W., Neshvizhskii, A., Ghosh, D., and Qin, Z. S. (2009). Hierarchical hidden markov model with application to joint analysis of ChIP-chip and ChIP-seq data. *Bioinformatics* **25**, 1715 – 1721.
114. Patwa, T. H., Li, C., Poisson, L. M., Pal, M., Ghosh, D., Simeone, D. M., and Lubman, D. M. (2009). The identification of phosphoglycerate kinase-1 and histone H4 autoantibodies in pancreatic cancer patient serum using a natural protein microarray. *Electrophoresis* **30**, 2215-2226.
115. Ghosh, D. and Chakrabarti, R. (2009). Joint classification and variable selection for immuno-histochemical data. *Biomarker Insights* **4**, 103 – 110. Available at the following URL: [http://la-press.com/article.php?article\\_id=1557](http://la-press.com/article.php?article_id=1557).
116. Khankal, R., Chin, J. W., Ghosh, D. and Cirino, P. C. (2009). Transcriptional effects of CRP\* expression in *E. coli*. *Journal of Biological Engineering* **3**, 13.
117. Gudjonsson, J. E., Ding, J., Li, X., Nair, R., Stuart, P., Tejasvi, T., Qin, Z. S., Ghosh, D., Aphale, A., Gumucio, D. L., Voorhees, J. J., Abecasis, G., Elder, J. T. (2009). Global gene expression analysis reveals evidence for decreased lipid biosynthesis and increased innate immunity in uninvolved psoriatic skin. *Journal of Investigative Dermatology* **129**, 2795 – 2804.
118. Brinkmeier, M. L., Davis, S. W., Carninci, P., MacDonald, J. W., Kawai, J., Ghosh, D., Hayashizaki, Y., Lyons, R. H., and Camper, S. A. (2009). Discovery of transcriptional regulators and signaling pathways in the developing pituitary gland by bioinformatic and genomic approaches. *Genomics* **93**, 449 – 460.
119. \*Chen, W., Ghosh, D., Raghunathan, T. E. and Sargent, D. L. (2009). Bayesian variable selection with joint modeling of categorical and survival outcomes: an application to individualizing chemotherapy treatment in advanced colorectal cancer. *Biometrics* **65**, 1030 – 1040. (**Note:** This paper won Wei Chen an SBSS student award for JSM 2007).

120. Ghosh, D. (2009). Assessing significance of peptide matches in proteomics experiments: a multiple testing approach, *Statistics in Biosciences* **1**, 199 – 213.
121. Ghosh, D. and Choi, H. W. (2009). Discussion of “A Bayesian model for cross-study differential gene expression” by Scharpf et al. *Journal of the American Statistical Association* **104**, 1310 – 1314.
122. \*Choi, H. W., Qin, Z. S. and Ghosh, D. (2010). A double-layered mixture model for the joint analysis of DNA copy number and gene expression data. *Journal of Computational Biology* **17**, 199 – 213.
123. Ghosh, D. (2010). Discrete nonparametric algorithms for outlier detection with genomic data. *Journal of Biopharmaceutical Statistics* **20**, 193 – 208.
124. Ghosh, D. (2010). Analysis of recurrent events: artificial censoring, truncation, pairwise estimation and inference, *Lifetime Data Analysis* **16**, 509 – 524.
125. Ghosh, D. and Qin, Z. S. (2010). Statistical issues in the analysis of ChIP-Seq and RNA-Seq data. *Genes* **1**, 317 – 334. Available at <http://www.mdpi.com/2073-4425/1/2/317/>
126. Ghosh, D., Elliott, M. R. and Taylor, J. M. G. (2010). Links between surrogate endpoints and endogeneity. *Statistics in Medicine* **29**, 2869 – 2879.
127. Ghosh, D. and Sabel, M. (2010). Spline-based modelling for predictiveness curves and surfaces. *Statistics and Its Interface* **3**, 445 – 454.
128. Ghosh, D. (2010). Detecting outlier genes from high-dimensional data: a fuzzy approach, *International Journal of Systems and Synthetic Biology* **1**, 273 – 283.
129. Ghosh, D. (2011). Propensity score modelling in observational studies using dimension reduction methods. *Statistics and Probability Letters* **81**, 813 – 820.
130. Ghosh, D. (2011). Generalized Benjamini-Hochberg procedures using spacings. *Journal of the Indian Society of Agricultural Statistics* **65**, 213 – 220.
131. Chen, F., Ding, X., Ding, Y., Xiang, Z., Ghosh, D., Schurig, G., Sriranganathan, N., Boyle, S., and He. Y. (2011). Proinflammatory caspase-2 mediated macrophage cell death induced by a rough attenuated *Brucella suis*. *Infection and Immunity* **79**, 2460 – 2469.
132. Camper, S., Mortensen, A., MacDonald, J. and Ghosh, D. (2011). Candidate genes for panhypopituitarism identified by gene expression profiling, *Physiological Genomics* **43**, 1105 – 16.
133. \*Poisson, L. M., Taylor, J. M. G. and Ghosh, D. (2011). Integrative set enrichment testing for multiple omics platforms. *BMC Bioinformatics* **12**, 459.
134. Chang, S., Tchitchek, N., Ghosh, D., Benecke, A. and Katze, M. G. (2011). Expression of a chemokine gene signature derived from meta-analysis predicts the pathogenicity of viral respiratory infections, *BMC Systems Biology* **5**, 202.

135. \*Li, Y. and Ghosh, D. (2012). Assumption weighting for incorporating heterogeneity into meta-analysis of genomic data, *Bioinformatics* **28**, 807 – 814.
136. Ghosh, D., Taylor, J. M. G. and Sargent, D. J. (2012). Meta-analysis for surrogacy: accelerated failure time modelling and semi-competing risks (with discussion), *Biometrics* **68**, 226 – 247.
137. Begum, F., Ghosh, D., Tseng, G. C. and Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Research* **40**, 3777 – 84.
138. Tseng, G. C., Ghosh, D. and Feingold, E. (2012). Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Research* **40**, 3785 – 99.
139. \*Poisson, L. M., Sreekumar, A., Chinnaiyan, A. M. and Ghosh, D. (2012). Pathway-based weighting strategies for combining gene expression and metabolomics data, *Genomics* **99**, 265 – 74.
140. Ghosh, D. (2012). A causal framework for surrogate endpoints with semi-competing risks data, *Statistics and Probability Letters* **82**, 1898-1902.
141. Chen, W., Ghosh, D., Raghunathan, T. E., Norkin, M., Sargent, D. J. and Bepler, G. (2012). On Bayesian methods of exploring qualitative interactions for targeted treatment. *Statistics in Medicine* **31**, 3693 – 3707.
142. Ghosh, D. (2012). Incorporating the empirical null hypothesis into the Benjamini-Hochberg procedure. *Statistical Applications in Genetics and Molecular Biology* **11**, 4.
143. Ghosh D. (2013). Genomic outlier detection in high-throughput data analysis. *Methods Mol Biol.* **972**, 141 – 53. doi: 10.1007/978-1-60327-337-4\_9. PubMed PMID: 23385536.
144. \*Choi, H., Fermin, D., Neshvizhskii, A., Ghosh, D. and Qin, Z. S. (2013). Sparsely correlated hidden Markov models with application to genome-wide location studies. *Bioinformatics* **29**, 533-541.
145. \*Lee, Y., Ghosh, D. and Zhang, Y. (2013). Association testing for gene-gene interactions on sex chromosomes, *Frontiers in Genetics* **4**, 239.
146. Teo, G., Kim, S., Ghosh, D., and Choi, H. W. (2014). Analysis of parallel gene and protein expression data for temporal cellular response to stress, *Journal of Proteomic Research* **13**, 29 – 37.
147. \*Lee, Y., Ghosh, D., Hardison, R. C. and Zhang, Y. (2014). MRHMMs: Multiple Regression Hidden Markov Models and other variants. *Bioinformatics* **30**, 1755 – 1756.
148. \*Lee, Y., Ghosh, D. and Zhang, Y. (2014). Regression hidden Markov modeling reveals heterogeneous gene expression regulation: a case study in mouse embryonic stem cells, *BMC Genomics* **15**, 360.

149. \*Li, Y. and Ghosh, D. (2014). A two-step hierarchical set testing framework, with applications to gene expression data on ordered categories, *BMC Bioinformatics* **15**, 108.
150. \*Phillips, D. and Ghosh, D. (2014). Testing the disjunction hypothesis using Voronoi diagrams with applications to genetics, *Annals of Applied Statistics* **8**, 801 – 823.
151. Ghosh, D. and Li, S. (2014). Unsupervised outlier profile analysis, *Cancer Informatics Suppl.* **4**, 25 – 33.
152. \*Li, Y. and Ghosh, D. (2014). Meta-analysis based on weighted ordered p-values for genomic data with heterogeneity. *BMC Bioinformatics* **15**, 226.
153. de Bekker, C., Quevillon, L., Smith, P. B., Fleming, K., Ghosh, D., Patterson, A. D., and Hughes, D. P. (2014). Species-specific ant brain manipulation by a specialized 1 fungal parasite. *BMC Evolutionary Biology* **14**, 166.
154. Ghosh, D. (2014). An asymptotically minimax kernel machine. *Statistics and Probability Letters*, 33 – 38.
155. \*Zhu, Y., Ghosh, D., Mitra, N., and Mukherjee, B. (2014). A data-adaptive strategy for inverse weighted estimation of causal effects. *Health Services Outcomes and Research Methodology* **14**, 69 – 91.
156. Almli, L. M., Duncan, R., Feng, H., Ghosh, D., Binder, E. B., Bradley, B., Ressler, K., Conneely, K. N. and Epstein, M. P. (2014). Resolving systematic bias in genetic association tests that consider interaction effects: application to a genome-wide association study of post-traumatic stress disorder. *JAMA Psychiatry* **71**, 1392 – 1399.
157. \*Hua, W. Y., Nichols, T. E. and Ghosh, D. for the Alzheimer's Disease Neuroimaging Initiative. (2015). Multiple comparison procedures for neuroimaging genomewide association studies. *Biostatistics* **16**, 17 – 30.
158. \*Zhan, X. and Ghosh, D. (2015). Incorporating auxiliary information for improved prediction using combinations of kernel machines. *Statistical Methodology* **22**, 47 – 57.
159. \*Zhu, Y., Coffman, D. L. and Ghosh, D. (2015). A boosting algorithm for estimating generalized propensity scores with continuous treatments. *Journal of Causal Inference* **3**, 25 – 40.
160. \*Zhan, X., Ghosh, D. and Epstein, M. P. (2015). Adaptive genetic association test using double kernel machines, *Statistics in Biosciences* **7**, 262 – 281.
161. Ghosh, D., Zhu, Y. and Coffman, D. L. (2015). Penalized regression procedures for variable selection in the potential outcomes framework. *Statistics in Medicine* **34**, 1645 – 58.
162. Ge, T., Nichols, T. E., Ghosh, D., Mormino, E. C., Smoller, J. W. and Sabuncu, M. R.; Alzheimer's Disease Neuroimaging Initiative. A kernel machine method for detecting effects of interaction between multidimensional variable sets: an imaging genetics application. *Neuroimage* 2015 Apr 1;109:505 – 14.

163. \*Zhan, X., Patterson, A. D. and Ghosh, D. (2015). Kernel approaches for differential analysis of mass spectrometry-based metabolomics data, *BMC Bioinformatics*, DOI: 10.1186/s12859-015-0506-3, URL: <http://www.biomedcentral.com/1471-2105/16/77>
164. \*Hua, W.-Y. and Ghosh, D. (2015), Equivalence of kernel machine regression and kernel distance covariance for multidimensional phenotype association studies. *Biometrics* **71**, 812 – 820.
165. \*Cho, Y. and Ghosh, D. (2015). Weighted estimation of the accelerated failure time model in the presence of dependent censoring. *PLoS One*. 2015 Apr 24;10(4): e0124381.
166. \*Zhu, Y., Ghosh, D., Coffman, D. L. and Savage, J. S. (2016). Estimating controlled direct effects of restrictive feeding practices in the "Early dieting in girls" study. *Journal of the Royal Statistical Society Series C* **65**, 115–130.
167. Traynor, A. J., Aragon, M., Ghosh, D., Choi, R. S., Dingmann, C., Tran, Z., and Bucklin, B. A. (2016). Obstetric Anesthesia Workforce Survey: 30-Year Update. *Anesth Analg*. 2016 Jun;122(6):1939-46.
168. Amini, A., Jones, B. T., McDermott, J., Serracino, H., Jimeno, A., Raben, D., Ghosh, D., Bowles, and Karam, S. (2016). Survival outcomes with concurrent chemoradiation for locally advanced elderly head and neck cancer patients using the National Cancer Data Base. *Cancer* **122**, 1533-43.
169. Broadaway, K. A., Cutler, D. J., Duncan, R., Moore, J. L., Ware, E. B., Jhun, M. A., Bielak, L. F., Zhao, W., Smith, J. A., Peyser, P. A., Kardia, S. L. R., Ghosh, D., and Epstein, M. P. (2016). A statistical approach for testing cross-phenotype effects of rare variants, *American Journal of Human Genetics* **98**, 525 – 540.
170. Ghosh, D. (2016). A modified risk set approach for biomarker evaluation studies, *Statistics in Biosciences* **8**, 395 – 406.
171. Waxman, T., Amini, A., Jones, B. L., McDermott, J. D., Raben, D., Ghosh, D., Bowles, D. W., and Karam, S. D. (2016). Association of Adjuvant Chemoradiotherapy vs Radiotherapy Alone With Survival in Patients With Resected Major Salivary Gland Carcinoma: Data From the National Cancer Data Base. *JAMA Otolaryngol Head Neck Surg* **142**, 1100 – 1110.
172. Zhan, X., Girirajan S., Zhao, N., Wu, M.C. and Ghosh, D. (2016). A novel kernel method for copy number variant association analysis with application to autism spectrum disorder, *Bioinformatics* **32**, 3603 – 3610.
173. \*Zhan, X. and Ghosh, D. (2016). A novel power-based approach to Gaussian kernel selection. *Statistical Methodology* **33**, 180 – 191.
174. Amini, A., Jones, B. T., Ghosh, D., Schechter, T. and Goodman, K. (2017). Impact of facility volume on outcomes in patients with squamous cell carcinoma of the anal canal: Analysis of the National Cancer Data Base. *Cancer* **123**, 228 – 236.



175. \*Cho, Y. and Ghosh, D. (2017). A general approach to goodness of fit for U-processes, *Statistica Sinica* **27**, 1175 – 1192.
176. Zhang, W., Epstein, M., Fingerlin, T. E. and Ghosh, D. (2017). Links between the sequence kernel association and kernel-based adaptive cluster tests, *Statistics in Biosciences* **9**, 246 – 258.
177. Luo, W., Zhu, Y. and Ghosh, D. (2017). On estimating regression causal effects using sufficient dimension reduction, *Biometrika* **104**, 51 – 65.
178. Stokes, W., Amini, A., Jones, B. L., McDermott, J. D., Raben, D., Ghosh, D., Bowles, D. W., and Karam, S. D. (2017) Induction Chemotherapy Offers No Overall Survival Benefit in Advanced Head and Neck Cancer: a National Cancer Data Base Analysis. *Head and Neck* **39**, 1113 – 1121.
179. Choi, H., Ghosh, D. and Qin, Z. (2017). Computationally Tractable Multivariate HMM in Genome-Wide Mapping Studies. *Methods Mol Biol* **1552**, 135 – 148.
180. Gorman, W. E., Kong, D. S., Balboni, I. M., Rudra, P., Bolen, C. R., Ghosh, D., Davis, M. M., Nolan, G. P., Hsieh, E. W. Y. (2017). Mass Cytometry Demonstrates a Distinct JAK/STAT dependent Monocyte Cytokine Signature Shared by Clinically Heterogeneous Pediatric SLE Patients. *Journal of Autoimmunity*, doi: 10.1016/j.jaut.2017.03.010.
181. Kim, D., Volk, H., Girirajan, S., Pendergrass, S., Hall, M. A., Verma, S. S., Schmidt, S. S., Hansen, R. L., Ghosh, D., Hertz-Picciotto, I., Ritchie, M.D. and Selleck, S. B. (2017). The joint effect of air pollution exposure and copy number variation on risk for autism. *Autism Research*, doi: 10.1002/aur.1799.
182. Duex, J. E., Owens, C., Chauca-Diaz, A., Dancik, G. M., Vanderlinden L. A., Ghosh, D., Leivo, M. Z., Hansel, D. E., Theodorescu, D. (2017). Nuclear CD24 drives tumor growth and is associated with metastatic disease and reduced survival in patients. *Cancer Research*, doi: 10.1158/0008-5472.CAN-17-0367.
183. Zhang, W. and Ghosh, D. (2017). On the use of kernel machine methods for Mendelian randomization studies. *Quantitative Biology* **5**, 368 – 379.
184. Lutz, S. M., Thwing, A., Schmiede, S., Kroehl, M., Baker, C., Starling, A., Ghosh, D. and Hokanson J. E. (2017). Examining the role of unmeasured confounding in mediation analysis with genetic and genomic applications, *BMC Bioinformatics* **18**: 344.
185. Amini A, Jones BL, Stumpf P, Leong S, Lieu CH, Weekes C, Davis SL, Messersmith WA, Purcell WT, Ghosh D, Schefter T, Goodman KA. Patterns of Care for Locally Advanced Pancreatic Adenocarcinoma Using the National Cancer Database. *Pancreas* 2017 Aug;46(7):904-912. doi: 10.1097/MPA.0000000000000876. PubMed PMID: 28697131.
186. Zhu, B., Song, N., Shen, R., Arora, A., Machiel, M. J., Song, L., Landi, M. T., Ghosh, D., Chatterjee, N., Balandandayuthapani, V., and Zhao, H. (2017). Integrating Clinical and Multiple Omics Data for Prognostic Assessments across Human Cancers, *Scientific Reports* **7**, 16954.
187. \*Cho, Y, Hu, C., and Ghosh, D. (2018) Covariate adjustment using propensity scores for

dependent censoring problems in the accelerated failure time model. *Statistics in Medicine* **37**, 390 – 404. <https://doi.org/10.1002/sim.7513>.

188. \*Philtron, D., Lyu, Y., Li, Q. and Ghosh, D. (2018). Maximum rank reproducibility: a non-parametric approach to assessing reproducibility in replicate high-throughput experiments, *JASA - Applications and Case Studies* 113, 1028-1039.

189. Zhang L, Zhou H, Li X, Vartuli RL, Rowse M, Xing Y, Rudra P, Ghosh D, Zhao R, Ford HL. Eya3 partners with PP2A to induce c-Myc stabilization and tumor progression. *Nat Commun*. 2018 Mar 13;9(1):1047. doi: 10.1038/s41467-018-03327-4. Erratum in: *Nat Commun*. 2018 Sep 17;9(1):3830. PubMed PMID: 29535359; PubMed Central PMCID: PMC5849647.

190. \*Rudra, P., Broadaway, K. A., Ware, E. B., Jhun, M. A., Bielak, L. F., Zhao, W., Smith, J. A., Peyser, P. A., Kardia, S. L. R., Epstein, M. P. and Ghosh, D. (2018). Testing cross-phenotype effects of rare variants in longitudinal studies of complex traits, *Genetic Epidemiology* 2018 Mar 30. doi: 10.1002/gepi.22121.

191. †Smith, D. E., Golden, A., Stange, B., Cragle, D., Barker, L., Morz, P., Weitzenkamp, D., Baron, A., Ghosh, D. and Newman, L. (2018). Beryllium lymphocyte proliferation test (Be-LPT) Uninterpretables: an assessment of overproliferation and poor growth results, *American Journal of Industrial Medicine* 2018 Mar 24. doi: 10.1002/ajim.22842.

192. Vartuli, R. L., Zhou, H., Zhang, L., Powers, R. K., Klarquist, J., Rudra, P., Vincent, M. Y., Ghosh, D., Costello, J. C., Kedl, R. M., Slansky, J. E., Zhao, R., and Ford, H. L. (2018). Eya3 Enhances Breast Cancer Progression via Threonine Phosphatase Mediated Upregulation of PD-L1 and Resultant Immune Suppression. *Journal of Clinical Investigation* 2018 May 14. pii: 96784. doi: 10.1172/JCI96784.

193. Zhu, Y., Savage, J. S. and Ghosh, D. (2018). A kernel-based diagnostic for balance assessment, *Journal of Causal Inference* 2018 Sep;6(2). pii: 20160029. Available at <https://www.degruyter.com/view/j/jci.ahead-of-print/jci-2016-0029/jci-2016-0029.xml>.

194. Russell, P., Fountain, K., Wolverton, D, and Ghosh, D. (2018). TCIPathfinder: an R client for the Cancer Imaging Archive REST API. *Cancer Research* **78**, 4424 – 26. Available at <https://www.biorxiv.org/content/early/2017/12/30/240986>.

195. Ghosh, D. (2018). Relaxed covariate overlap and margin-based causal effect estimation, *Statistics in Medicine*. 2018 Dec 10;37(28):4252-4265. doi: 10.1002/sim.7919. Epub 2018 Aug 30. PubMed PMID: 30168168. Available at <https://arxiv.org/abs/1801.00816>.

196. Marzec, L., Raghavan, S., Banaei-kashani, F., Creasy, S., Melanson, E. L., Lange, L., Ghosh, D. and Rosenberg, M. A. (2018). Device-Measured Physical Activity Data for Classification of Patients with Ventricular Arrhythmia Events: A Pilot Investigation, *PLoS One* 2018 Oct 29;13(10):e0206153.

197. Wylie, K. P., Harris, J. G., Ghosh, D., Olincy, A. and Tregellas, J. R. (2019). MATRICS Consensus Working Memory is associated with distributed Executive Control Networks during resting state fMRI. *Journal of Neuropsychiatry and Clinical Neurosciences*. 2019 May 23:ap-

pineuropsych18060131. doi: 10.1176/appi.neuropsych.18060131. [Epub ahead of print] PubMed PMID: 31117908.

198. Bennett, T. D., Callahan, T., Feinstein, J. A., Ghosh, D., Lakhani, S. A., Spaeder, M. D., Szefer, S. J., and Kahn, M. G. (2019). Data science for child health, *Journal of Pediatrics* **208**, 12 – 22.

199. King, J., Russell, S., Bennett, T. D. and Ghosh, D. (2019). Kung Faux Pandas: Simplifying privacy protection. *AMIA Jt Summits Transl Sci Proc.* 2019 May 6;2019:267-274. eCollection 2019. PubMed PMID: 31258979. Available at <https://openreview.net/forum?id=BJe0v23m1X>

200. Oliphant, M.U.J., Vincent, M.Y., Galbraith, M.D., Pandey, A., Zaberezhnyy, V., Rudra, P., Johnson K.R., Costello, J.C., Ghosh, D., DeGregori, J.V., Espinosa, J.M., and Ford, H.L. (2019). Six2 regulates late-stage metastasis via direct regulation of Sox2 and induction of a cancer stem cell program, *Cancer Research*, in press.

201. Vestal, B. E., Carlson, N. E., Estepar, R. S.-J., Fingerlin, T., Ghosh, D., Kechris, K. and Lynch, D. (2019). Using a spatial point process framework to characterize lung computed tomography scans, *Spatial Statistics* **29**, 243 – 267.

202. Russell, P. H. and Ghosh, D. (2019). Radtools: R utilities for smooth navigation of medical image data. 2018 Dec 24;7. pii: ISCB Comm J-1976. doi: 10.12688/f1000research.17139.1. eCollection 2018. PubMed PMID: 31131079; PubMed Central PMCID: PMC6518432.

203. Russell S, Bennett TD, Ghosh D. (2019). Software engineering principles to improve quality and performance of R software. *PeerJ Comput. Sci.* 5:e175 DOI 10.7717/peerj-cs.175.

204. Ozdemir, T., Bowers, D., Zhan, X., Ghosh, D., Polson, S., and Brown, J. L. (2019). High-throughput siRNA screening reveals signaling pathways connecting substrate curvature guided cell shape and osteoinduction. *Scientific Reports.* 2019 Jan 30;9(1):1001.

205. Xing, F., Cornish, T., Bennett, T. D., Ghosh, D. and Yang, L. (2019). Pixel-to-pixel Learning with Weak Supervision for Single-stage Nucleus Recognition in Ki67 Images, *IEEE Transactions in Biomedical Engineering.*

206. \*Jensen, A., Tregellas, J., Sutton, B., Xing, F. and Ghosh, D. (2019). Flexible kernel machine-based association tests between brain networks and phenotypes, *PLoS One* **14**, e0199340. Available at <https://www.biorxiv.org/content/10.1101/341495v1>.

207. Holleman, A. M., Broadaway, K. A., Duncan, R., Almlı, L. M., Bradley, B., Ressler, K. J., Ghosh, D., Mülle, J. G., and Epstein, M. P. (2019). GAMBITS: An improved method for genetic association testing of symptom and questionnaire data in psychiatric genetic studies, *Scientific Reports*, in press. Available at <https://www.biorxiv.org/content/early/2018/08/03/383471>.

208. Rao, S., Ghosh, D., Asturias, E. J. and Weinberg, A. (2019). What can we learn about influenza infection and vaccination from transcriptomics? *Human Vaccines and Immunotherapeutics* **22**, 1 – 9. doi: 10.1080/21645515.2019.1608744. [Epub ahead of print] PubMed PMID: 31116679.

209. Anantharajan, J., Zhou, H., Zhang, L., Hotz, T., Vincent, M., Blevins, M., Jones, D., Jasson, A. E., Kuan, J. W. L., Ng, E. Y., Khoon, Y. Y., Baburajendran, N., Lin, G., Hung, A. W., Joy, J., Patnaik, S., Marugan, J., Rudra, P., Ghosh, D., Hill, J., Keller, T. H., Zhao, R., Ford, H. and Kang, C. (2019). Structural and functional analyses of an allosteric Eya2 phosphatase inhibitor, *Molecular Cancer Therapeutics* **18**, 1484 – 1496.
210. Ghosh, D. (2019). Wavelet-based Benjamini-Hochberg procedures for multiple testing under dependence, *Mathematical Biosciences and Engineering* **17**, 56 – 72.
211. Saltzman AF, Smith, DE, Gao D, Ghosh D, Amini A, Aldrink JH, Dasgupta R, Gow KW, Glick RD, Ehrlich PF and Cost NG. (2019). How Many Lymph Nodes Are Enough? Assessing the Adequacy of Lymph Node Yield for Staging in Favorable Histology Wilms Tumor. *Journal of Pediatric Surgery* **54**, 2331 – 2335.
212. Ghosh, D. and Cruz-Cortés, E. (2019). A Gaussian process framework for overlap and causal effect estimation with high-dimensional covariates, *Journal of Causal Inference* **7**, 2. Available at <http://arxiv.org/abs/1801.03185>.
213. Schuyler, R., Jackson, C., Garcia-Perez, J. E., Baxter, R. M., Ogolla, S., Rochford, R., Ghosh, D., Rudra, P., and Hsieh, E. W. Y. (2019). Minimizing batch effects in mass cytometry data, *Frontiers in Immunology* **10**, 2367.
214. Xing F, Bennett T, Ghosh D. Adversarial Domain Adaptation and Pseudo-Labeling for Cross-Modality Microscopy Image Quantification. *Med Image Comput Comput Assist Interv.* 2019 Oct;11764:740-749.
215. Major, A., Smith, D., Ghosh, D., Rabinovitch, R. and Kamdar, M. (2020). Risk and types of secondary primary malignancies in DLBCL survivors changes over time based on stage at diagnosis, *Cancer* **126**, 189 – 201.
216. Solis-Lemus, C. R., Fischer, S. T., Liu, C., Leslie, E. J., Cutler, D. J., Ghosh, D., and Epstein, M. P. (2020). Enhanced case-control association analyses leveraging family history, *Genetics*, accepted.
217. Tiwari, P., Colborn, K., Smith, D. E., Xing, F., Ghosh, D. and Rosenberg, M. A. (2020). Application of Deep-Learning to Prediction of Incident Atrial Fibrillation using Harmonized Electronic Health Record Data. *JAMA Network Open* **3**, e1919396. Available at <https://www.biorxiv.org/content/10.1101/520866v1>.
218. \*Rudra, P., \*Cruz-Cortés, E., Zhang, X. and Ghosh, D. (2020). Multiple testing approaches for hypotheses in integrative genomics, *Wiley Interdisciplinary Reviews in Computational Statistics*, in press.
219. \*Cruz-Cortés, E.. and Ghosh, D. (2020). An invitation to system-wide algorithmic fairness, accepted. In *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society*.
220. Shaikh, S. S., Kamath, S. D., Ghosh, D., Lewandowski, R. J. and McMahon, B. J. (2020). Safety and outcomes of permanent and retrievable inferior vena cava filters in the oncology popula-

tion, *International Journal of Vascular Medicine*. Available at <https://www.hindawi.com/journals/ijvm/2020/6582742/>

221. \*Josey, K. P., Juárez-Colunga, E., Yang, F. and Ghosh, D. (2020). A framework for covariate balance using Bregman divergences. *Scandinavian Journal of Statistics*, accepted. Available at <https://arxiv.org/pdf/1903.00390.pdf>

222. Warsavage, T., Xing, F., Barón, A. E., Feser, W. J., Hirsch, E., Miller, Y., Malkoski, S. Wolf, H. J., Wilson, D. O. and Ghosh, D. (2020). Quantifying the incremental value of deep learning: Application to lung nodule detection. *PLoS One*. 2020 Apr 14;15(4):e0231468. doi: 10.1371/journal.pone.0231468. PMID: 32287288; PMCID: PMC7156089.

223. Czaja, A. S., Collins, K., Valuck, R. J., Anderson, H. D., Ghosh, D. and Davidson, J. (2020). Validity of administrative claims-based algorithms for ventricular arrhythmia and cardiac arrest in the pediatric population. *Pharmacoepidemiology Drug Safety*. 2020 Nov;29(11):1499-1503. doi: 10.1002/pds.5001. Epub 2020 Apr 13. PMID: 32283564.

224. \*Zhang, X., Cornish, T. C., Yang, L., Bennett, T. D., Ghosh, D. and Xing, F. (2020). Generative Adversarial Domain Adaptation for Nucleus Quantification in Ki67 Images. *Journal of Clinical Oncology - Cancer Clinical Informatics* 4, 666 – 679. doi:10.1200/CCI.19.00108

225. Czaja, A. S., Anderson, H., Ghosh, D., Davidson, J., Campbell, J. and Valuck, R. (2020). Selective serotonin reuptake inhibitor use is associated with increased ventricular arrhythmias in children and young adults: a case-control study. *Journal of Pediatrics* **226**, 173-178.

226. Lorberbaum, D. S., Kishore, S., Rosselet, C., Sarbaugh, D., Aragon, E., Xuan, S., Simon, O., Ghosh, D., Mendelsohn, C., Gadue, P., Sussel, L. (2021). Retinoic acid signaling within pancreatic endocrine progenitors regulates mouse and human pancreas expression. *Development*, in press.

227. \*Cho, Y., and Ghosh, D. (2021). Covariate adjustment for recurrent events with dependent censoring, *Communications in Statistics: Theory and Methods* **50**, 216 – 236.

228. Okamoto, Y., Ghosh, T., Okamoto, T., Schuyler, R., Seifert, J., August, J., Lenis-Charry, L., Visser, A., Fleischer, C., Feser, M., Bemis, E. A., Norris, J. M., Kuhn, K. A., Demoruelle, M. K., Deane, K. D., Ghosh, D., Holers, V. M., Hsieh, E. W. Y. (2021). Mass cytometry identifies enhanced histone H3 citrullination and pro-inflammatory cytokine production by toll-like receptor stimulated CD14<sup>high</sup> monocytes in subjects at-risk for future development of rheumatoid arthritis, *J Autoimmunity*, in press.

229. F. Xing, T. C. Cornish, T. D. Bennett and D. Ghosh. (2021). Bidirectional Mapping-Based Domain Adaptation for Nucleus Detection in Cross-Modality Microscopy Images. *IEEE Transactions on Medical Imaging*, doi: 10.1109/TMI.2020.3042789.

230. Cho, Y. and Ghosh, D. (2021). Quantile-based subgroup identification in clinical trials. *Statistics in Biosciences* **13**, 90 – 128.

231. Zhang, W. and Ghosh, D. (2021). A general sensitivity analysis approach for Mendelian randomization. *Statistics in Biosciences* **13**, 34 – 55.

232. Vestal, B. E., Carlson, N. E. and Ghosh, D. (2021). Filtering spatial point patterns using kernel densities. *Spatial Statistics*, in press.
233. \*Josey, K. P., Ghosh, D., Berkowitz, S. and Raghavan, S. (2021). Extending inferences of experimental results with entropy balancing, *Statistics in Medicine*, in press.
234. Cho, Y., Hu, C. and Ghosh, D. (2021) Analysis of regression discontinuity designs using censored data, *Journal of Statistical Research* **55**, 225 – 248. Available at <https://arxiv.org/pdf/1908.03646.pdf>.
235. Buchwald AG, Bayham J, Adams J, Bortz D, Colborn K, Zarella O, Buran M, Samet J, Ghosh D, Herlihy R, Carlton EJ. Estimating the Impact of Statewide Policies to Reduce Spread of Severe Acute Respiratory Syndrome Coronavirus 2 in Real Time, Colorado, USA. *Emerg Infect Dis*. 2021 Jun 30;27(9). doi: 10.3201/eid2709.204167.
236. Bear, J., Sargent, J., O'Neill, B., Chapman, K., Ghosh, D., Kirsch, H. and Tregellas, J. (2021). Spike-associated networks predict postsurgical outcomes in children with refractory epilepsy, *Journal of Clinical Neurophysiology*, accepted.
237. Raghavan, S., Josey, K., Bahn, G., Berkowitz, S. A., Basu, S., Reaven, P., Hayward, R., and Ghosh, D. (2021). Generalizability of heterogeneous treatment effects of intensive glycemic control on all-cause mortality across two randomized clinical trials, *Annals of Epidemiology*, accepted.
238. Ghosh, T., Philtrou, D., Zhang, W., Kechris, K. and Ghosh, D. (2021). Reproducibility of mass spectrometry-based metabolomics data, *BMC Bioinformatics*, accepted.
239. Ghosh, D. and Sabel, M. (2021). A weighted sample framework to incorporating external calculators for risk prediction, *Statistics in Biosciences*, accepted.
240. Simon, O., Buard, I., Rojas, D. C., Holden, S. K., Kluger, B. M., and Ghosh, D. (2021) A Novel Approach to Understanding Parkinsonian Cognitive Decline Using Minimum Spanning Trees, Edge Cutting, and Magnetoencephalography, *Scientific Reports*, accepted.
241. Dekermanjian, J., Labeikovsky, W., Ghosh, D. and Kechris, K. (2021). MSCAT: A Machine Learning assisted Catalog of Metabolomics Software Tools, *Metabolites*, available at <https://www.mdpi.com/2218-1989/11/10/678>.
242. Iain R. Konigsberg, Bret Barnes, Monica Campbell, Elizabeth Davidson, Yingfei Zhen, Olivia Pallisard, Meher Boorgula, Corey Cox, Debmalya Nandy, Souvik Seal, Kristy Crooks, Evan Sticca, Genelle F. Harrison, Andrew Hopkinson, Alexis Vest, Cosby G. Arnold, Michael G. Kahn, David P. Kao, Brett R. Peterson, Stephen J. Wicks, Debashis Ghosh, Steve Horvath, Wanding Zhou, Rasika A. Mathias, Paul J. Norman, Rishi Porecha, Ivana V. Yang, Christopher R. Gignoux, Andrew Monte, Alem Taye, Kathleen C. Barnes. (2021). Host methylation predicts SARS-CoV-2 infection and clinical outcome, *NPJ Communications Medicine*, accepted.
243. Carpenter, C. M., Zhang, W, Severn, C., Gillenwater, L., Ghosh, T., Kechris, K., Ghosh, D. (2021). PaIRKAT: A pathway integrated regression-based kernel association test with applications to lung function data, *PLoS Computational Biology*, accepted.

244. Cruz-Cortés, E., Rajtmajer, S. M. and Ghosh, D. (2021). Structural Interventions on Automated Decision Making Systems. In *Proceedings of the NeurIPS workshop on Algorithmic Fairness through the Lens of Causality and Robustness*, in press.

245. Zhang, Y., Xing, F., Ghosh, D., Banaei-Kashani, F., Bowler, R. P. and Kechris, K. (2021). Development of an augmented high-dimensional Graphical Lasso model to incorporate prior biological knowledge for global network learning, *Frontiers in Genetics*, in press.

246. Johnson, M. J., Liu, C., Ghosh, D., Lang, N., Levin, M., and Weinberg, A. (2021) Cell-mediated immune responses after administration of the recombinant zoster vaccine: five-year resistance, *Journal of Infectious Diseases*, in press.

247. Rudra, P., Baxter, R. P., Hsieh, E. W. Y, and Ghosh, D. (2021). Compositional data analysis using kernels in mass cytometry data, *Bioinformatics Advances*, accepted.

248. Simon, O. B., Rojas, D. C., Ghosh, D., Yang, X., Rogers, S., Martin, C. S., Holden, S. K., Kluger, B. M., and Buard, I. (2021) Profiling Parkinsons disease cognitive phenotypes via resting-state magnetoencephalography, *Journal of Neurophysiology*, accepted.

#### *Submitted/Under Revision*

249. Dekermanjian, J., Shaddox, E., Nandy, D., Ghosh, D. and Kechris, K. (2022). Mechanism-Aware Imputation: A two-step approach in handling missing values in metabolomics, under revision.

250. Huang, Y., Zhu, Y., Cai, X., and Ghosh, D. (2022). Causal mediation analysis based on partial linear structural equation models, under revision.

251. Severn, C., Suresh, K, Görg, C., Choi, Y. S., Jain, R. and Ghosh, D. (2022). A pipeline for the implementation and visualization of explainable machine learning in medical imaging, in preparation.

252. Josey, K. P., Yang, F., Ghosh, D., and Raghavan, S. (2022). A calibration approach to transportability with observational data, under revision.

253. Vu, T., Wrobel, J., Bitler, B. G., Schenk, E., Jordan, K. and Ghosh, D. (2022). SPF: A spatial and functional data analytic approach to cell imaging data, submitted.

254. Simon, O. B., Jain, R., Choi, Y. S., Goerg, C., Suresh, K., Severn, C. and Ghosh, D. (2022). A unified approach to the analysis of radiomics of glioma using minimum spanning trees, under revision.

255. Edward, J. A., Josey, K., Bahn, G., Caplan, L., Reusch, J. E. B., Reaven, P., Ghosh, D. and Raghavan, S. (2022). Heterogeneous treatment effects of intensive glycemic control on major adverse cardiovascular events in the ACCORD and VADT trials: a machine learning analysis, submitted.

256. Vestal, B. E., Ghosh, D., Estepar, R. S.-J., Fingerlin, T., Kechris, K. and Carlson, N. E. (2022).

Quantifying the spatial clustering characteristics of radiologically-based emphysema, submitted.

257. Ghosh, D., Amini, A., Jones, B. and Karam, S. D. (2022). Population-level and individual-level explainers for propensity score matching in observational studies, under revision.

258. Ghosh, D., Josey, K. P., Cruz-Cortés, E. and Yang, F. (2022). A framework for evaluating matching procedures, submitted.

260. Cho, Y., Zhan, X. and Ghosh, D. (2022). Nonlinear predictive directions in clinical trials, under revision.

261. Seal, S., Vu, T., Ghosh, T., Wrobel, J., and Ghosh, D. (2022). DenVar: density-based variation analysis of multiplex imaging data, submitted.

262. Ghosh, D., Mastej, E., Jain, R. and Choi, Y. (2022). Causal models in radiomics: framework, mechanisms and algorithms, submitted.

263. Ghosh, D. (2022). Sufficient dimension reduction: an information-theoretic viewpoint, under revision.

264. Suresh, K., Severn, C. and Ghosh, D. (2022). Survival prediction models: an introduction to discrete-time modeling, submitted.

265. Seal, S., Wrobel, J., Johnson, A. M., Nemenoff, R. A., Schenk, E. L., Bitler, B. G., Jordan, K. R., and Ghosh, D. (2022). On clustering for cell phenotyping in multiplex immunohistochemistry (mIHC) and multiplexed ion beam imaging (MIBI) Data, submitted.

#### *In preparation*

Ghosh, D. (2022) On optimal slicing for sufficient dimension reduction, in preparation.

Zuo, S., Yang, F., Josey, K. P., Raghavan, S. and Ghosh, D. (2022). Transporting clinical trial results with time-to-event data in colorectal cancer, a case study.

Ghosh, D., Cruz-Cortés, E. and Raghavan, S. (2022). Overlap assumptions and margin-based inference for transportable causal effects, in preparation.

Cho, Y. and Ghosh, D. (2022) Some nonlinear methods for Mendelian randomization, in preparation.

Ghosh, D. (2021). Prediction-aware biomarker evaluation, in preparation.

#### **Publications which have not received peer review:**

Ghosh, M. and Ghosh, D. (2021). Comment on 'Bayesian Restricted Likelihood Methods: Conditioning on Insufficient Statistics in Bayesian Regression' by Lewis et al., *Bayesian Analysis*, to appear.



Cruz-Cortés, E., Yang, F., Juaréz-Colunga, E., Warsavage, T. and Ghosh, D. (2021). Comment on Statistical Modelling: the Two Cultures by Leo Breiman. *Observational Studies*, in press.

Ghosh, D. and Cho, Y. (2020). Bridging linearity-based and kernel-based sufficient dimension reduction. Available at <https://arxiv.org/abs/2010.15009>.

Ghosh, D. and Cho, Y. (2020). Predictive directions for individualized treatment selection in clinical trials. Available at <https://arxiv.org/abs/1807.03375>.

Bennett, T. D., Russell, S., King, J., Schilling, L. D., Voong, C., Rogers, N., Adrian, B., Bruce, N., and Ghosh, D. (2019). Sepsis detection by the Epic sepsis model versus an early warning score. Available at <http://arxiv.org/abs/1902.07276>.

Smith, K., Ghosh, D., Pollard, K. A., and De, S. (2018). A computational framework for detecting signatures of accelerated somatic evolution in cancer genomes. Available at <https://www.biorxiv.org/content/early/2017/08/16/177261>

Kechris, K. and Ghosh, D. (2017). Discussion on ‘Statistical contributions to bioinformatics: design, modelling, structure learning and integration’ by J. S. Morris and V. Balandandayuthapani. *Statistical Modelling* **17**, 327 – 331.

Daniels, M., Frangakis, C., Charu, V. and Ghosh, D. (2015). University of Pennsylvania seventh annual conference on statistical issues in clinical trials: Current issues regarding the use of biomarkers and surrogate endpoints in clinical trials (morning panel discussion). *Clin Trials*. 2015 Aug;12(4):323-32.

Ghosh, D. (2008). Progression-free survival as a surrogate for overall survival: a new paradigm for surrogacy?. Technical Report, Department of Statistics, Penn State University.

Ghosh, D. and Chinnaiyan, A. M. (2005). Statistical methods for identification of tumor progression genes in microarray data. Technical Report, Department of Biostatistics, University of Michigan.

Ghosh, D. (2005). Semiparametric methods for the binormal model with multiple biomarkers. Technical Report, Department of Biostatistics, University of Michigan.

Shen, R., Ghosh, D. and Chinnaiyan, A. M. (2004). A two-stage mixture model for meta-analysis of microarray data. Technical Report, Department of Biostatistics, University of Michigan.

Ghosh, D. (2004). Model checking techniques for regression models in cancer screening. Technical Report, Department of Biostatistics, University of Michigan.

Chen, W., Ghosh, D., Raghunathan, T. E. and Kardia, S. (2004). A Bayesian method for selection of genomic interactions. Technical Report, Department of Biostatistics, University of Michigan. (**Note:** This paper won Wei Chen an ENAR student travel award in 2003).

## **Chapters of books**

Ghosh T, Zhang W, Ghosh D, Kechris K. Predictive Modeling for Metabolomics Data. *Methods*

Mol Biol. 2020;2104:313-336. doi: 10.1007/978-1-0716-0239-3\_16. PMID: 31953824; PMCID: PMC7423323.

Reinhold, D., Pielke-Lombardo, H., Jacobson, S., Ghosh, D. and Kechris, K. (2018). Pre-analytic considerations for mass spectrometry based untargeted metabolomics data. In A. D'Alessandro (Ed.), *Methods for Molecular Biology*. Methods Mol Biol. 2019;1978:323-340. doi: 10.1007/978-1-4939-9236-2\_20. PubMed PMID: 31119672.

Coffman, D. L., MacKinnon, D. P., Zhu, Y., and Ghosh, D. (2016). A comparison of potential outcomes approaches for assessing causal mediation. In H. He, P. Wu, and D.-G. Chen (Eds.), *Statistical Causal Inferences and Their Applications in Public Health Research*. New York: Springer.

Ghosh, D. and Girirajan, S. (2014). Statistical considerations in the analysis of rare variant data. In *Statistical Analysis of Next-generation Sequencing Data*, ed: S. Datta and D. Nettleton. New York: Springer-Verlag, Chapter 20.

Poisson, L. M. and Ghosh, D. (2013). Functional enrichment methods: a review. In *Advances in Statistical Bioinformatics*, ed: Kim-Anh Do, Steve Qin, and Marina Vannucci. Cambridge University Press: Cambridge. Chapter 21.

Ghosh, D., Lee, S., and Park, T. (2010). Statistical models, inference and algorithms for large-scale biological data analysis. In *Statistical Bioinformatics*, ed: Jae K. Lee. New York: Wiley-Blackwell: 185 – 199.

Ghosh, D. (2009). Oncogenic outlier detection in high-throughput data. In *Statistical Methods for Microarray Data Analysis*, eds: A. Yakovlev, D. Gaile, and L. Klebanov. Humana Press.

Ghosh, D., Rhodes, D. and Chinnaiyan, A. M. (2009). Combining genomic data in human studies. In *Meta-analysis and Combining Information in Genetics and Genomics*, eds: D. B. Allison and R. Guerra. Chapman and Hall/CRC Mathematical and Computational Biology, in press.

Ghosh, D. (2008). Multiple testing procedures under confounding. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in honor of Prof. Pranab K. Sen*, eds: M. Silvapulle, E. Pena and N. Balakrishnan. IMS Monograph Series: 243 – 256.

Choi, H. and Ghosh, D. (2008). A comparison of meta-analysis methods for gene expression data. In *Statistical Advances in the Biomedical Sciences*, eds: A. Biswas, S. Datta, J. P. Fine and M. R. Segal. Wiley: 200 - 215.

Ghosh, D. (2000). Response conditional models for correlated binary data. In *Perspectives in Statistical Science*, ed. A.K. Basu *et al.* New Delhi: Oxford University Press, pp. 138–152.

### **Nonrefereed Publications:**

Ghosh, D. (2021). Review of *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences* by Gramacy. *International Statistical Review*, to appear.

Ghosh, D. (2021). Review of *Statistical Inference Via Convex Optimization* by Juditsky and Ne-

mirovskii, *International Statistical Review*, to appear.

Ghosh, D. (2017). Review of *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry* by Datta and Mertens, *International Statistical Review*, to appear.

Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM. (2010). Re: Florian Jentzmik, Carsten Stephan, Kurt Miller, et al. Sarcosine in Urine after Digital Rectal Examination Fails as a Marker in Prostate Cancer Detection and Identification of Aggressive Tumours. *Eur Urol*.

Ghosh, D. (2009). Review of *Modern Multivariate Statistical Techniques* by A. Izenman. *Biometrics* **65**, 990 – 991

Ghosh, D. (2009). Letter to the Editor on “Effect of preventive supplementation with ready-to-use therapeutic food on the nutritional status, mortality, and morbidity of children aged 6 to 60 months in Niger: a cluster randomized trial,” *Journal of the American Medical Association* **301**, 2208 – 2209.

Wang, X., Ghosh, D. and Chinnaiyan, A. M. (2005). Response to “Autoantibody signatures in prostate cancer.” *New England Journal of Medicine* **26**, 2816 – 2817.

Ghosh, D. (2002). Review of *Statistical Aspects of the Design and Analysis of Clinical Trials*, by B. S. Everitt and A. Pickles. *Controlled Clinical Trials* **23**, 299 – 300.

Ghosh, D. (2001). Review of *Post-genome informatics*, by M. Kanehisa. *Statistical Methods in Medical Research* **10**, 241–242.

Ghosh, D. (2001). Review of *Analysis of Multivariate Survival Data*, by P. Hougaard. *Statistical Methods in Medical Research* **10**, 306–307.

## **Current Grants**

NSF DMS-1914937 (Ghosh, P.I.) 9/1/19 - 8/31/22  
New methods in high-dimensional causal inference  
Total \$50,000.00/year

In this proposal, the PI and team will focus their research in two areas. The first will be to understand the implications of deep learning algorithms and their performance on foundational assumptions for the popular potential outcomes model. The second part of the grant will deal with developing scalable algorithms for causal effect estimation. New computationally scalable algorithms for causal effect estimation will be developed as part of this proposal.

Role: Principal Investigator

NIH U01CA235488-01 (MPIs: K. Kechris-Mays, D Ghosh), 8/15/18 - 7/31/22  
Addressing Sparsity in Metabolomics Data Analysis  
Total: \$275,000/year

In this proposal, we will consider sparsity that occurs in a variety of steps in the metabolomics

preprocessing pipeline: i.e, in preprocessing, normalization, differential expression and pathway analysis. We will develop new statistical and computational tools to handle sparsity that occurs in the various steps of metabolomics data analysis.

Role: Principal Investigator

NIH R01CA129102 (MPIs: J. M. G. Taylor, University of Michigan; D. Ghosh), 01/01/09 - 12/31/2022

Statistical Methods for Cancer Biomarkers

Total: \$100,000/year (UC-Denver subcontract)

This grant proposes new semiparametric and nonparametric modelling procedures with applications to cancer biomarkers. The aims are the following: (a) Development of semiparametric and nonparametric multivariate isotonic regression modelling procedures for biomarkers; (b) Development of statistical methods for the analysis of surrogate endpoints in a single-trial and multiple-trial framework; (c) Development of hybrid model averaging methods and attendant projection-based framework for combining biomarkers to optimize predictive accuracy.

Role: Principal Investigator (PI of UC-Denver subcontract)

NIH R01 CA221282 (Ford/Zhao, MPI) 07/01/2017-06/30/2022

Role of Eya3 in regulating the immune microenvironment to promote breast tumor progression

In this proposal we will test the hypothesis that Eya3 regulates breast tumor growth and progression via its ability to recruit PP2A, leading to increased levels of PD-L1 and a diminished tumor-specific T-cell response.

Role: Biostatistician

NIH R01 CA224867 (Ford) 08/01/2018 - 05/31/2022

Examining the EYA2/MYC axis in Group 3 Medulloblastoma

Work in this proposal will test a new model for medulloblastomas and will test the roles of EYA2 and Myc in a subpopulation of medulloblastomas.

Role: Biostatistician

NIH U01 AI141919 (Weinberg) 01/01/2019-12/31/2023

Persistence of Protection Conferred by Shingrix Against Herpes Zoster in Older Adults

This proposal aims to study local and systemic transcriptomic and epigenetic changes in response to shingrix vaccination with two different types of vaccines.

Role: co-Investigator

NIH R01 HL142049 (Carlson/Maier/Fingerlin) 04/01/2019 03/31/2023

Novel integrative approaches for disease phenotyping, utilizing radiomics in Sarcoidosis The goals of this proposal are to develop reproducible radiographic phenotypes of pulmonary sarcoidosis and integrate radiographic data with clinical data, genetic variants and transcriptional signatures, re-defining sarcoidosis biomarkers.

Role: Co-Investigator

## **Completed Funding**

R21 CA237493 (Xing) 04/01/2019-03/31/2021

NIH/NCI \$275,000/2 years

Development and Dissemination of KiNet: a novel imaging informatics tool for gastrointestinal and pancreatic neuroendocrine tumors In this proposed research, we seek to develop and disseminate

a novel deep learning-based imaging informatics system, KiNeT, specifically for better automated Ki67 LI measurement in GI and pancreatic NETs. It will use deep fully convolutional networks to develop an end-to-end, pixel-to-pixel model for single-stage Ki67 LI assessment.

Role: co-Investigator

NIH R01GM117946 (MPIs: M. Epstein, Emory University; D Ghosh) 1/1/16 - 12/31/19

Statistical Tests for Mapping Genetic Determinants of Complex Traits

Total \$125,000/year (UC-Denver subcontract)

Genome-wide association studies (GWAS) and next-generation sequencing (NGS) projects have uncovered only a limited number of trait-influencing loci. While large increases in sample size will improve power to detect such variation, the ascertainment and sequencing/genotyping of such samples are costly and inefficient. Therefore, it is desirable to increase power to detect such variants without requiring additional sample collection. We propose novel methods for improved gene mapping of common and rare susceptibility variants that move beyond standard strategies typically applied to GWAS and NGS studies of complex traits.

Role: Principal Investigator (PI of UC-Denver subcontract)

NSF ABI-1262538 (Ghosh, P.I.) 7/1/13 - 3/31/17

Multivariate Statistical Methods for Genomic Data Integration

Total \$177,027.00/year

This research focuses on multivariate methods of analysis with high-dimensional genomic data, with the goal of prioritizing the genome for further study. Two classes of problems will be studied during the course of the project. The first is Hidden Markov Models and the second is multiple testing procedures, whose use have become commonplace with genomic datasets. This project proposes novel multivariate extensions of both types of method with a goal of being characterized by sound theoretical statistical principles while simultaneously being computationally feasible on big datasets. The methodology will be evaluated using several real datasets as well as through simulation studies.

Role: Principal Investigator

NIH 1T32GM102057-01A1 (MPIs: R. Hardison, D. Ghosh, and C. Shashikant, Penn State University), 07/01/13 - 06/30/14

Total: \$80,713/Year 1

Computation, Bioinformatics, and Statistics (CBIOS) Training Program

Genomic data are transforming how scientists in medicine and basic science conduct research. The advancement of genome science requires a new generation of scientists with strong computational and statistical skills and the ability to effectively interact with experimentalists. The proposed Penn State Computation, Bioinformatics, and Statistics (CBIOS) Training Program will prepare a cadre of investigators to think innovatively and keep pace with the quickly evolving landscape of high throughput genomic technologies. The program faculty are interdisciplinary and highly collaborative, with expertise in computation, bioinformatics, statistics, functional, medical, and evolutionary genomics. Learning these discipline-crossing skills will make trainees competitive for future careers in emerging and rapidly advancing fields of comparative, systems, statistical and medical genomics.

Role: Co-Director/Principal Investigator

NIH UL1RR033184 (PI: L. Sinoway, Penn State University/Penn State College of Medicine), 09/01/11 - 06/30/2014

Penn State Clinical and Translational Institute

Total: \$5,468,625/year

This proposal describes the newly created Penn State Clinical and Translational Science Institute (CTSI). The overarching goal of our CTSI initiative is to revitalize the health science research and education enterprise at our University to better enable it to deliver on the promise of improved health.

Role: Biostatistician

NIH R01GM066411-05A2 Omiecinski (PI) 12/15/09 – 11/30/2013

Functional Analysis of Nuclear Receptor Variants

Total: \$314,000/year

In this grant, our studies will characterize the roles of a novel series of receptor proteins that interact with pharmaceuticals and environmental chemicals and function to regulate the liver's capacity to metabolize substances. These receptors are termed constitutive androstane receptors and are part of a battery of sensing proteins that exist with liver cells to assist in processing the nature of the cell's chemical environment so that it can better tune its ability to detoxify chemicals we are exposed to.

Role: Co-Investigator

NIH/NSF 1 R01 GM72007 (PI: D. Ghosh), 09/01/04 – 08/31/10

Statistical Methods for the Analysis of Functional Genomic Data

Total: \$150,000/year

The goal of this project is to develop statistical and bioinformatic procedures for the modeling of complex high dimensional biological data with an emphasis towards incorporating functional biological knowledge.

Role: Principal Investigator

Co-Investigators: A. Chinnaiyan (Univ. of Michigan, 2004-2007), F. Pugh (Penn State, 2007 - 2008)

NIH 5R01HG003618 (PI: M. Epstein, Emory University), 09/27/07 – 08/26/2009

Novel Statistical Methods for Human Gene Mapping

Total: \$200,000/year

This grant proposes a set of such statistical methods that either address novel problems or improve existing solutions to problems in human gene mapping studies. These proposed methods are applicable to a variety of genetic studies as they address topics in linkage, linkage disequilibrium, and high-dimensional genetic analyses of complex diseases and disease-related quantitative traits.

Role: Co-investigator

Collaborators: M. Epstein (Emory University)

NIH/NCRR M01 RR00042 (Omenn), 03/01/01 – 02/28/06

General Clinical Research Center

Total: \$145,749

The major goals of this project are to review protocols and consult on biostatistical and bioinformatic aspects of investigations that use the General Clinical Research Center.

American Cancer Society RSG-02-179-MGO (Chinnaiyan), 07/01/02 – 06/30/06

Molecular Classification of Prostate Cancer

Total: \$37,792.33

The major goal of this project is to advise researchers on the design, analysis and conduct of high-throughput studies in prostate cancer involving gene and protein expression microarrays.

NIH/NHLBI P30 CA46592 (Wicha), 06/01/99 – 05/31/04

Cancer Center Support Grant (Biostatistics Core)

Total: \$69,323

The major goal of this project is to provide statistical support to University of Michigan Cancer Center researchers in the areas of design, analysis and interpretation of data.

MUNN IDEA Grant (Ghosh)

7/1/02 – 6/30/03

Total: \$20,000

The goal was to develop Bayesian and penalized regression models for the classification of tumors.

Prostate SPORE Seed Grant (Ghosh)

7/1/02 – 6/30/03

Total: \$10,000

The goal was to develop Bayesian and penalized regression models for the classification of tumors.

NIH P60 DK20572 (Herman), 12/01/02 to 11/30/07

Michigan Diabetes Research and Training Center

Total: \$ 92,498.76

The major goal of this project is to advise diabetes researchers on statistical issues related to the use of microarrays.

NIH R03 A1054406A (Zhang), 04/01/03 to 03/31/05

Library on a Chip: Bacterial Strain Microarray

Total: \$5,636

The major goal of this project is to array the genomes of a library of bacterial strains on a single chip that can be probed for the presence or absence of specific genes and for the allelic variations of the genes.

NIH P50 CA069568 (Pienta), 04/01/03 to 05/31/08

SPORE in Prostate Cancer

Total: \$45,450

The major goal of this project is to provide statistical support to University of Michigan Cancer Center researchers in the areas of design, analysis and interpretation of data in prostate cancer studies.

## **Invited Talks**

1. "Navigating spatially-resolved cell imaging data":
  - Department of Biostatistics, Memorial Sloan Kettering, November 3, 2021.
2. "Introduction to Causal modelling in biomedical research":
  - Western North American Region Outreach Seminar Series, July 16, 2021.
3. "Enhancing interpretability of machine learning tools in radiomics":
  - Joint Statistical Meetings, August 5, 2020.
  - Western North American Region Meeting, June 15, 2021.
4. "Surprises in high-dimensional causal inference":
  - Department of Biostatistics, University of Michigan, November 14, 2019.
  - Biostatistics Group, Division of Cancer Epidemiology and Genetics, National Cancer Institute, September 7, 2020.
  - Quantitative Sciences Group, Department of Medicine, Stanford University, November 11, 2020.
5. "Evaluating reproducibility with high-dimensional data"
  - Department of Computational Medicine and Biology, University of Michigan, Wednesday, November 13, 2019.
6. "Revisiting propensity scores":
  - University of Colorado Data Science to Patient Value (D2V) seminar, May 21, 2019.
7. "Sufficient dimension reduction and covariate overlap in causal inference":
  - 10th International Triennial Calcutta Symposium in Statistics, Calcutta University, December 28, 2018.
  - Department of Biostatistics, Mailman School of Public Health, Columbia University, November 14, 2018.
8. "Thinking causally with high-dimensional databases":
  - Institute for Mathematics and its Applications, Minneapolis, Minnesota, November 6, 2018.
  - Canadian Statistical Science Institute-National Institute of Statistical Science Health Data Science Workshop, May 7, 2021.
9. "Building multi-scale kernels for fMRI data using the Morlet transform"
  - Joint Statistical Meetings, Baltimore, Maryland, August 1, 2017.
10. "Outlier profile analysis, with applications to cancer":
  - Biostatistical Modelling: A Conference in Honor of Jeremy Taylor's 60th Birthday, Ann Arbor, MI, June 10, 2017.



11. "Model selection and estimation in causal inference":
  - Division of Biostatistics, University of Indiana, Indianapolis, IN, April 21, 2017.
  - Division of Biostatistics, Ohio State University, Columbus, OH, April 14, 2017.
12. "Big data: what are they and where are we headed with them":
  - Keynote Speaker, ACCORDS Workshop on Big Data, Anschutz Medical Campus, October 13, 2016.
13. "Some applications of machine learning methods to causal inference":
  - Keynote Session, International Indian Statistical Association, Corvallis, OR, August 20, 2016.
  - Division of Biostatistics, University of Minnesota, Minneapolis, MN, April 6, 2016.
  - Statistics Research Seminar, Department of Mathematical and Statistical Sciences, UC Denver, Denver, CO, March 29, 2016.
  - Division of Biostatistics, University of Miami, Miami, FL, October 20, 2015.
14. "Kernel machine methods: back to the future":
  - Colloquium, Department of Statistics, Colorado State University, Ft. Collins, CO, December 5, 2016.
  - Myrto Lefkopolou Distinguished Lecture, Department of Biostatistics, Harvard School of Public Health, Boston, MA, September 24, 2015.
15. "Stacking, support vector machines and censored data":
  - Graybill/International Chinese Student Association Conference, Ft. Collins, Colorado, June 16, 2015.
16. "Funding for NIH grants: a reviewer's perspective":
  - ENAR Junior Researchers Workshop, Miami, Florida, March 15, 2015.
  - Graybill/International Chinese Student Association Conference, Ft. Collins, Colorado, June 16, 2015.
17. "Kernel machine methods for high-throughput data":
  - Keynote Speaker, Ohio Mass Spectrometry and Metabolomics Symposium, Columbus, Ohio, May 17, 2018.
  - Computational Biosciences Seminar Series, University of Colorado Health Sciences Center, Aurora, Colorado, February 2, 2015.
  - Beyond Bioinformatics Workshop, Statistical and Applied Mathematical Sciences Institute, September 14, 2014.
18. "Measurement, inference and statistical science in the age of 'Big Data'":
  - American Public Health Association Annual Meeting, Boston, MA, November 5, 2013.
19. "Multivariate statistical methods for genomic data integration":

- Department of Statistics, University of Pennsylvania, Philadelphia, PA, March 27, 2014.
  - Department of Biostatistics and Medical Informatics, University of Wisconsin, January 7, 2014.
  - Department of Biostatistics and Informatics, University of Colorado, December 16, 2013.
  - Department of Biostatistics, University of Florida, November 14, 2013.
  - Joint seminar, Department of Biostatistics and Division of Oncology Biostatistics, Johns Hopkins University, August 28, 2013.
20. “Penalized regression methods for variable selection in causal inference”:
- Joint Statistical Meetings, Montreal, Canada, August 5, 2013.
21. “Data-adaptive modelling of propensity scores for causal inference”:
- Division of Biostatistics, Penn State College of Medicine, Hershey, PA, November 15, 2012.
22. “Kernel Machines: an overview and recent developments”:
- Department of Biostatistics, University of Michigan, Ann Arbor, MI, November 1, 2012.
  - Department of Biostatistics, M. D. Anderson Cancer Center, Houston, TX, February 19, 2013.
23. “Multivariate multiple testing procedures, with applications to finding gene fusions in TCGA data”:
- Joint Statistical Meetings, San Diego, CA, August 1, 2012.
24. “Introducing shrinkage into the Benjamini-Hochberg procedure”:
- International Chinese Statistical Association Applied Conference, Boston, MA, June 24, 2012.
25. “Multiple testing procedures in neuroimaging genomics”:
- ENAR Annual Spring Meeting, Washington, DC, April 4, 2012.
26. “Extending the Benjamini-Hochberg procedure using spacings”:
- Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, January 19, 2012.
  - Department of Statistics, University of South Carolina, Columbia, SC, February 1, 2012.
  - Department of Biostatistics, University of North Carolina, Chapel Hill, NC, February 22, 2012.
  - Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD, March 13, 2012.
27. “Meta-analysis of genomic data and multiple testing”:
- Cancer Biostatistics Seminar, Department of Biostatistics, University of Michigan, Ann Arbor, MI, November 2, 2012.

- Starr Lectureship, Department of Biostatistics, Brown University, Providence, RI, November 14, 2011.
  - Division of Biostatistics, Vanderbilt University, Nashville, TN, November 16, 2011.
28. "A multivariate Benjamini-Hochberg procedure, with applications to genomics":
- Indian International Statistical Association meeting, Raleigh, NC, March 24, 2011.
29. "What can machine learning do for you? Algorithmic approaches to causal inference":
- Methodology Center, Brownbag Seminar, Penn State University, February 21, 2011.
30. "Risk prediction: model averaging, stability and calibration":
- Division of Oncology Biostatistics, Johns Hopkins University, Baltimore, MD, February 17, 2011.
31. "Multiple testing: a view using spacings":
- Winter Workshop on high-dimensional data analysis, University of Florida, Gainesville, FL, January 14, 2011.
32. "Identification of copy number-associated gene expression alterations in microarray experiments":
- Joint Statistical Meetings, Washington, D. C., August 4, 2009.
33. "Kernel machine-based methods in genomics":
- Rao Prize Conference, Penn State University, University Park, PA, May 22, 2009.
34. "Model-based meta-analysis for analysis of data from chromatin immunoprecipitation experiments":
- Division of Biostatistics, University of Southern California, Los Angeles, CA, July 23, 2009.
  - IMS Asia/Pacific Rim Meeting, Seoul, South Korea, June 29, 2009.
  - Department of Biostatistics and Bioinformatics and Human Genetics, Emory University, Atlanta, GA, April 23, 2009.
  - Division of Statistical Sciences, Cornell University, Ithaca, NY, April 1, 2009.
35. "Hierarchical Hidden Markov model-based meta-analysis of data from chromatin immunoprecipitation experiments":
- Department of Human Genetics, UCLA, Los Angeles, CA, October 13, 2008.
36. "Genomic outlier profile analysis with estimated null distributions":
- Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, September 26, 2008.
37. "Statistical issues in the design and analysis of '-omics' studies":

- Bioinformatics and Medical Informatics Retreat, Penn State College of Medicine, Hershey, PA, January 22, 2008.
38. "Multiple testing procedures in genomic studies":
- Silver Platinum Jubilee, Indian Statistical Institute, Calcutta, India, January 3, 2008.
39. "Kernel machines, mixed models and genomic studies":
- Division of Biostatistics, Yale University, New Haven, CT, September 25, 2007.
  - Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, PA, September 19, 2007.
  - Department of Biological Statistics and Computational Biology, Cornell University, September 12, 2007.
40. "Statistical methods for the analysis of phage epitope array data":
- Statistics for Biomolecular Data Integration and Modeling Workshop, Ascona, Switzerland, June 11, 2007.
  - Interface meeting, Philadelphia, PA, May 25, 2007.
41. "Integrating copy number and gene expression microarray datasets":
- Department of Epidemiology and Biostatistics, Memorial Sloan Kettering, New York City, New York, May 9, 2007.
42. "Towards inference of chromosomal aberrations from genomic data":
- Department of Bioinformatics and Computational Biology, M. D. Anderson Cancer Center, Houston, Texas, October 2, 2006.
43. "Joint modelling of copy number and mRNA microarray data":
- 2007 ENAR annual meeting, Atlanta, Georgia, March 12, 2007.
44. "Statistical Methods for Integration of Copy Number and transcript mRNA data":
- Department of Statistics, Rice University, Houston, Texas, February 5, 2007.
  - Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, D. C., September 1, 2006.
  - Joint Statistical Meetings, Seattle, Washington, August 8, 2006.
45. "Statistical Methods for Analysis of Genomic/Proteomic Data in Complex Tissue":
- Statistical Society of Canada, University of Western Ontario, London, Ontario, May 29, 2006.
46. "Meta-analysis of genomic data and multiple testing":
- Department of Statistics, Pennsylvania State University, State College, PA, March 14, 2006.
  - Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, March 9, 2006.

47. "Introduction to Statistical Methods for Genomic Data Analysis":
  - Association of Molecular Pathology Annual Meeting, Phoenix, Arizona, November 11, 2005.
48. "Multiple testing and shrinkage estimation":
  - Multiple Comparisons Procedures, Shanghai, China, August 18, 2005.
49. "Towards inference of chromosomal aberrations from genomic data":
  - Joint Statistical Meetings, Minneapolis, Minnesota, August 10, 2005.
50. "Combining Genomic Data in Human Cancer Studies":
  - Division of Oncology Biostatistics, Department of Oncology, Johns Hopkins University, Baltimore, Maryland, March 1, 2006.
  - Mathematical Biosciences Institute, Columbus, Ohio, April 22, 2005.
51. "Semiparametric Support Vector Machines for Gene Expression Data":
  - Joint Statistical Meetings, Toronto, Canada, August 8, 2004.
  - Institute for Pure and Applied Mathematics, UCLA, June 4, 2004.
52. "Association Models for Bivariate Censored Data":
  - Division of Biostatistics, Cleveland Clinic, April 23, 2004.
  - Division of Biostatistics, Columbia University, March 4, 2004.
53. "Statistical Methods for Chromosomal Localization using Gene Expression Data":
  - Department of Biostatistics, University of Pittsburgh, February 17, 2005.
  - Division of Biostatistics, University of Minnesota, March 22, 2004.
  - Division of Biostatistics, Columbia University, March 3, 2004.
  - Department of Statistics, University of Tennessee-Knoxville, January 23, 2004.
  - Statistical Methods in Bioinformatics Seminar Series, Brown University, November 10, 2003.
54. "Statistical Methods for Clustering Microarray Data in Cancer Studies":
  - First Canadian Workshop on Statistical Genomics, Toronto, Canada, September 3, 2003.
55. "Statistical Methods for the Analysis of Microarray Data":
  - East Tennessee State University, January 22, 2004.
56. "Global cross ratio models for bivariate censored data":
  - Fifth Triennial Symposium on Statistics, Calcutta University, Calcutta, India, December 29, 2003.
  - Department of Biostatistics, Johns Hopkins University, April 30, 2003.

57. "Meta-analysis of microarray data":

- Bioinformatics Seminar Series, Department of Statistics, Purdue University, February 1, 2005.
- Biostatistics Seminar Series, Department of Statistics, University of Wisconsin-Madison, November 12, 2004.
- International Indian Statistical Association Meeting, Dekalb, IL, June 14, 2002.
- Joint Statistical Meetings, New York City, NY, August 8, 2002.

58. "Penalized Regression Models for the classification of tumors from microarray experiments":

- Center for Molecular Medicine and Genetics, Wayne State University, February 20, 2003.
- Institute for Pure and Applied Mathematics, UCLA, June 21, 2002.
- Department of Biostatistics, University of Minnesota, May 18, 2002.
- Training Program in Bioinformatics, Texas A&M University, February 25, 2002.
- Pacific Symposium on Biocomputing, Kauai, Hawaii, January 4, 2002.

59. "Mixture modelling of microarray data," Statistics Department, University of Washington, July 13, 2001.

60. "Semiparametric Analysis of Recurrent Failure Time Data Using Accelerated Rates Models":

- Department of Epidemiology and Biostatistics, Imperial Cancer Research Fund, London, England, June 9, 2002.
- Statistics 2001 Canada Conference, Concordia University, Montreal, Canada, July 2, 2001.

61. "Semiparametric Analysis of Recurrent Failure Time Data and Dependent Censoring":

- Joint Statistical Meetings, Atlanta, Georgia, August 7, 2001.
- Fourth Triennial Symposium on Statistics, Calcutta University, Calcutta, India, December 27, 2000.

62. "Marginal Regression Models for Recurrent and Terminal Events":

- Department of Biostatistics, M.D. Anderson Cancer Center, Houston, Texas, March 27, 2000.
- Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, South Carolina, March 24, 2000.
- Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, March 9, 2000.
- Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, March 6, 2000.
- Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, March 3, 2000.
- Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, February 24, 2000.

- Department of Statistics, University of Florida, Gainesville, Florida, February 17, 2000.
63. "Nonparametric Analysis of Recurrent Events and Death," Society of Clinical Trials Annual Meeting, Anaheim, California, May 3, 1999.
  64. "Response Conditional Models for Correlated Binary Data," Third Triennial Symposium on Statistics, Calcutta University, Calcutta, India, December 27, 1997.

### **Papers Delivered at Professional Meetings**

1. "Machine Learning Methods for Causal Inference," Contributed Paper, ENAR Spring Meeting, 2006, Tampa, Florida.
2. "Semiparametric Analysis of Recurrent Events and Dependent Censoring," Contributed Paper, ENAR Spring Meeting, 2001, Charlotte, North Carolina.
3. "Nonparametric Analysis of Recurrent Events and Death," Contributed Paper, ENAR Spring Meeting, 1999, Atlanta, Georgia.

### **Teaching:**

#### **University of Michigan:**

Biostatistics 646 (Data Analysis in Molecular Biology): Winter 2007  
 Biostatistics 699 (Design and Analysis of Biomedical Investigations): Winter 2007  
 Biostatistics 503 (Applied Biostatistics): Fall 2001  
 Biostatistics 560 (Statistical Methods in Epidemiology): Fall 2002, Fall 2003, Fall 2004, Fall 2005  
 Biostatistics 602 (Biostatistical Inference): Winter 2002, Winter 2003, Winter 2005  
 Biostatistics 830 (Machine Learning Methods in Biostatistics): Fall 2005  
 Guest Lecturer in Bioinformatics 526: Fall 2002, Fall 2003  
 Guest Lecturer in Biostatistics 646 (Data Analysis in Molecular Biology): Winter 2005  
 University of Michigan Cancer Bioinformatics Short Course Instructor: Summer 2004  
 Guest Lecturer in Human Genetics 632: Fall 2006  
 Guest Lecturer in Epidemiology 631: Winter 2007

#### **Penn State University:**

Statistics 525 - Survival Analysis: Spring 2008, Spring 2009  
 Statistics 505 - Applied Multivariate Statistical Analysis: Fall 2009, Fall 2010, Fall 2011  
 Statistics 544 - Categorical Data Analysis: Fall 2010  
 Instructor, Summer School for Astronomers: Summer 2009, Summer 2010, Summer 2011

#### **University of Colorado:**

BIOS 6640 - R and Python for Data Science: Spring 2016  
 BIOS 6640 - R for Data Science: Fall 2019

BIOS 6641 - Causal Analytics in Public Health: Spring 2017  
Guest Lecturer in Computational Biosciences 7711 (Intro to Bioinformatics I): Fall 2015 (1 lecture),  
Fall 2016 (1 lecture), Fall 2017 (2 lectures), Fall 2018 (1 lecture)  
Guest Lecturer in Computational Biosciences 7712 (Intro to Bioinformatics II): Spring 2016 (2 lec-  
tures), Spring 2017 (2 lectures), Spring 2018 (2 lectures)

## **Service**

### **University of Michigan:**

Faculty Search: 2000-2001, 2001-2002, 2005-2006  
Candidacy: 2000-2001, 2001-2002, 2003 - 2004, 2005-2006 (chair), 2006 – 2007 (chair)  
Statistical Genetics Search Committee: 2001-2002, 2002-2003  
Web: 2002-2003  
Associate Director, Cancer Biostatistics Training Grant: 2005 – 2006  
Faculty-Student Affairs: 2003 - 2004  
Microarray Working Group Organizer: 2001 - 2004  
Bioinformatics Faculty Search Committee: 2001-2002  
Aging/Bioinformatics Faculty Search Committee: 2003-2004  
Bioinformatics Executive Advisory Committee: 2004 – 2005  
Bioinformatics Graduate Affairs Committee: 2006 – 2007  
Chair, Bioinformatics Pilot Grant Award Program: 2006  
Genome Sciences Training Grant Faculty: 2001-2007  
Genome Sciences Training Grant Review Panel: 2002, 2004, 2007

### **Penn State University:**

Eberly Chair Search Committee: 2007  
Cliff Clogg Memorial Lecture Committee: 2008  
Ph.D. Qualifying Exam Committee: 2008, 2009  
Promotion and Tenure Committee: 2010  
Graduate Student Affairs Committee: 2009, 2010, 2011  
Admissions Committee: 2010, 2011  
Statistics Department Head Search Committee: 2009  
Associate Dean Search Committee: 2011  
CBIOS Training Grant Co-Director, 2013 – 2014

### **University of Colorado:**

Chair, Space Committee, Colorado School of Public Health, 2014 – 2017  
Member, Space Committee, Colorado School of Public Health, 2017 – present  
Internal Reviewer, Department of Environmental and Occupational Health, , Colorado School of  
Public Health, 2015  
Chair, Search Committee, Associate Dean for Finance and Administration, Colorado School of  
Public Health, 2015  
Search Committee, Molecular Diagnostics Laboratory Co-Director, Colorado Center for Personal-



ized Medicine, 2015

Search Committee, Medical Genetics Faculty Position, Colorado Center for Personalized Medicine, 2015

Investigator-Initiated Trial Review Committee, University of Colorado Cancer Center, 2016 – present  
Internal Advisory Board, University of Colorado Lung SPORE, 2017

**External:**

Statistical Reviewer, *Radiology*, 2015 – 2017

Referee for the following journals:

*Biometrics, Statistics in Medicine Journal of Multivariate Analysis, Biostatistics, Statistics and Probability Letters, Lifetime Data Analysis, Bioinformatics, Biometrika, Genome Biology, Journal of the American Statistical Association - Theory and Methods, Journal of Computational and Graphical Statistics, IMA Volume Series, Bayesian Statistics 7, Nature Medicine, Diabetes Care, Journal of Statistical Planning and Inference, Journal of Biological Chemistry, Journal of the American Statistical Association - Applications and Case Studies, Comptes rendus biologies, Neurocomputing, Journal of Multivariate Analysis, American Journal of Pharmacogenomics, Computational Statistics and Data Analysis, FEBS Letters, Annals of Statistics, Aging Cell, Comparative and Functional Genomics, Statistical Applications in Genetics and Molecular Biology, New England Journal of Medicine, IEEE Transactions in Computational Biology and Bioinformatics, Journal of Clinical Oncology, Journal of Very Large Databases, Journal of National Cancer Institute, BMC Bioinformatics, Applied Statistics, Lancet, Cancer Epidemiology, Biomarkers and Prevention, Journal of the Royal Statistical Society Series B, PLOS Genetics, Nucleic Acids Research, American Journal of Management Science, Journal of Computational and Applied Mathematics, Journal of Clinical Endocrinology and Metabolism, Statistical Advances in the Biomedical Sciences (book, 2 articles), Meta-Analysis in Genetics (book), Statistica Sinica, Journal of Molecular Diagnostics, Proceedings of the National Academy of Sciences, BMC Developmental Biology, Journal of Proteome Research, Cancer Informatics, Canadian Journal of Statistics, Mammalian Genome, Pacific Symposium on Biocomputing, International Journal of Biostatistics, Gastroenterology, Statistical Papers, Genomics, BMC Cancer, Current Molecular Medicine, Statistical Analysis and Data Mining, Statistics and Computing, Neuroimage, Annals of Epidemiology, Journal of Biopharmaceutical Statistics, American Journal of Epidemiology, Nature, American Statistician, Journal of Proteomics and Bioinformatics, Nature Medicine, TEST, Journal of Psychiatric Research, JAMA, PLOS One, Journal of Biomedical Informatics, Clinical Cancer Research, Cancer Research, Computational and Mathematical Methods in Medicine, Genetic Epidemiology, Genome Medicine, Trends in Analytical Chemistry, BMC Medicine, Molecular Psychiatry, Respiriology, Stat*

Invited Session Organizer, 2002 IISA Conference.

IMS Invited Session Organizer, 2003 Spring ENAR meeting.

NSF Grant Reviewer, 2003, 2005.

Book Reviewer, 2003.

Regional Advisory Board Member, ENAR, 2004 - 2006.

European Union Young Research Investigator Award Reviewer, 2004.

Medical Research Council *ad hoc* Grant Reviewer, 2004.

National Institutes of Health *ad hoc* Grant Reviewer, 2004, 2006, 2007, 2008, 2009 (2 times),

2015, 2017

National Institutes of Health Challenge Grant Reviewer, 2009.

NSA *ad hoc* Grant Reviewer, 2006.

University of Cyprus Grant Reviewer, 2006.

Reviewer, Springer-Verlag proposal, 2006.

Utah State University New Faculty Research Grant Reviewer, 2007.

Pennsylvania Department of Health Grant Reviewer, 2007.

Singapore Ministry of Health Reviewer, 2007.

Qatar National Government Grant Reviewer, 2007.

Distinguished Student Paper Award Committee, ENAR, 2007 - 2009, 2015 – 2016.

Panelist, Emerging Design and Analysis Issues in Genomic Studies in Population Sciences, Harvard School of Public Health, 2008.

Reviewer, Neural Information Processing Systems (NIPS) Conference, 2009.

Reviewer, ASA Statistics in Epidemiology Student Paper Competition, 2009.

U.S.-Israel Binational Science Foundation reviewer, 2010

Netherlands Organization for Health Research and Development reviewer, 2010.

Regular Member, Biostatistical Methods and Research Design (BMRD) Section, National Institutes of Health, 2010 – 2013

Member, Regional Committee (RECOM) board, ENAR, 2011 – 2013

Secretary/Treasurer, ASA Biometrics Section, 2011 – 2013

Reviewer, Feasibility Studies for Collaborative Interaction for Minority Institution/Cancer Center Partnership (P20), National Institutes of Health, 2011

Reviewer, The Netherlands Organisation for Health Research and Development, 2012

Reviewer, Utilizing the PLCO Biospecimens Resource to Bridge Gaps in Cancer Etiology and Early Detection Research (U01), 2013

External Reviewer, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 2014

Section Chair, ASA Biometrics Section, 2015 – 2016.

Organizing Committee, Frontiers of Hierarchical Modeling in Observational Studies, Complex Surveys and Big Data, 2014.

Reviewer, Israel Science Foundation, 2015, 2016.

Reviewer, NIH New Innovator's Award, 2016.

Reviewer, Precision Medicine Initiative Cohort Program Coordinating Center, 2016.

Discussant, Annual Health Econometrics Workshop, 2016.

Reviewer, BD2K R25 grants, 2016.

Short Course Instructor, "Machine Learning for Big Data using Python and Spark", 2nd Seattle Symposium on Healthcare Data Analytics, 2016.

Chair, NIH Special Emphasis Panel, Methodology and Measurement, 2017 (twice)

Co-Chair, National Institute of Environmental Health Sciences Special Emphasis Panel, Powering Research through Innovative Methods in Epidemiology, 2017.

Lester Curtin Award Committee, American Statistical Association, 2017.

Reviewer, NIH TOPMed Systems Biology Data Analysis RFA, 2017

Search Committee, *Biometrics* Co-Editor (European Region), 2018.

Reviewer, Cutting Edge Informatics Tools for Illuminating the Druggable Genome (U01) RFA, 2018.

Reviewer, Traceback Testing: Identification and Genetic Counseling of Mutation Carriers (U01) RFA, 2018.

Reviewer, Mechanisms of Disparities in Etiology and Outcomes of Lung Cancer in the U.S.: The Role of Risk and Protective Factors RFA, 2019.

Chair, Special Emphasis Panel for K awards, National Institutes of Environmental Health Sciences, 2019.

Panelist, Institute of Mathematical Statistics New Researchers Conference, 2019.

Panelist, CCTSI Summit on Artificial Intelligence and Machine Learning, 2019.

Reviewer, National Institutes of Environmental Health Sciences Career Mentored Awards, 2020.

Chair, NIH Special Emphasis Panel, Emergency Awards: Rapid Investigation of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Coronavirus Disease 2019 (COVID-19), 2020.

Panelist, Houston Chapter of the American Statistical Association Workshop on Communication, May 8, 2021.

Panelist, Frontiers in Causal Inference in Data Science: Perspectives from Leaders in Tech and Academia, University of Pennsylvania, May 28, 2021.

Reviewer, Program Project Grants (P01), National Cancer Institute, June 10-11, 2021.

Scientific Advisory Board, BEAMS (Binational Early Asthma and Microbiome Study) study, University of Arizona, 2021 – 2025.

Reviewer, National Science Foundation, 2021.

Program Committee, Society of Mathematical Biology Annual Meeting, 2004.

Program Committee, NSF-DARPA workshop on Genomic Signal Processing, 2002, 2005, 2006.

Program Committee, Intelligent Systems for Molecular Biology, 2005.

Program Committee (Member At-Large), ENAR, 2007.

Program Chair, Biometrics Section, JSM, 2008.

Program Committee, International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), 2011, 2014.

Program Chair, ENAR, 2012.

Program Committee, International Biometric Conference, 2014.

Program Committee, Graybill Conference, 2017.

Co-Chair, Program Committee, Biostatistical Modelling: A Conference in Honor of Jeremy Taylor's 60th Birthday, 2017.

Program Committee, 4th International Conference on Algorithms for Computational Biology, 2017.

Chair, Biometrics Section, American Statistical Association, 2016.

Chair, Section in Statistical Genetics and Genomics, American Statistical Association, 2018.

Associate Editor, *Biometrics*, 2006 - 2010, 2011 – 2017

Associate Editor, *Statistics in Biosciences*, 2009 – 2020

Editorial Board, *Molecular Cancer*, 2009 – 2017

Editorial Board, *Cancer Informatics*, 2005 – 2017

Associate Editor, *Statistica Sinica*, 2011 – 2014

Associate Editor, *BMC Bioinformatics*, 2009 - 2014

Associate Editor, *BMC Medical Genomics*, 2008 - 2014

Editorial Board, *BMC series*, 2005 – 2014

Editorial Board, *BMC Proceedings*, 2007 – 2014

Associate Editor, *EURASIP Journal on Signal Processing and Bioinformatics*, 2005 - 2008

Associate Editor, *JASA - Theory and Methods*, 2008 – 2011

Section Head, Bioinformatics and Algorithms, Section, *BMC Medical Genomics*, 2010 – 2012

Associate Editor, *Statistical Applications in Genetics and Molecular Biology*, 2009 – 2012

Associate Editor, *International Journal of Biostatistics*, 2009 – 2012  
Associate Editor, *Electronic Journal of Statistics*, 2022 – 2024

Member, American Statistical Association (1997 to present)  
Member, International Biometric Society (1998 to present)  
Member, Institute of Mathematical Statistics (1999 to present)  
Member, International Society for Computational Biology (2001 to 2003)  
Member, International Indian Statistical Association, 2008 - present

## **Student Advising**

### **Master's Students:**

#### **University of Michigan:**

Erin Shellman, Biostatistics, 2007

#### **Penn State University:**

Qiong Yang, Statistics, 2014

#### **University of Colorado:**

Katherine Roberts, Biostatistics and Informatics, 2015 – 2017

Alexandria Jensen, Biostatistics and Informatics, 2015 – 2017

Derek Smith, Biostatistics and Informatics, (joint with Anna Barón), 2016 – 2017

Ted Warsavage, Biostatistics and Informatics, (joint with Fuyong Xing and Anna Barón), 2017 – 2019

Cameron Severn, Biostatistics and Informatics, (joint with Krithika Suresh), 2019 – 2020

Jonathan Dekermanjian, Biostatistics and Informatics, (joint with Katerina Kechris), 2020 – 2021

### **Doctoral Students:**

#### **University of Michigan:**

Dawei Liu, Biostatistics, 2002 – 2005 (co-chair with X. Lin), currently Assistant Professor of Biostatistics, University of Iowa

Wei Chen, Biostatistics, 2002 – 2006 (co-chair with T.E. Raghunathan), currently Assistant Professor of Oncology, Wayne State University

Zheng Yuan, Biostatistics, 2004 – 2006, currently Ph.D. Biostatistician, Clinical Oncology, Eli Lilly

Ronglai Shen, Biostatistics, 2002 – 2007 (co-chair with J. Taylor), currently investigator, Division of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center

Hyungwon Choi, Biostatistics, 2006 – 2008 (co-chair with Z. S. Qin), currently assistant professor, Department of Biostatistics and Epidemiology, National University of Singapore Graduate Medical School

Laila Poisson, Biostatistics, 2003 – 2009 (co-chair with J. Taylor), current position, assistant professor, Department of Epidemiology and Biostatistics, Henry Ford Health System

**Penn State University:**

Muhammad Atiyat, Statistics, 2008 – 2011, currently Statistician, United Nations  
Scott Roths, Statistics, 2009 – 2011 (joint with Prof. G. Jogesh Babu), currently Lecturer,  
Department of Statistics, Penn State University  
Yeying Zhu, Statistics, 2010 – 2013 currently Assistant Professor,  
Department of Statistics and Actuarial Sciences, University of Waterloo  
Yihan Li, Statistics, 2010 – 2014, currently Biostatistician, AbbVie Pharmaceuticals  
Daisy Phillips, Statistics, 2011 – 2014, currently assistant professor  
Department of Statistics, Penn State University  
Wen-Yu Hua, Statistics, 2011 – 2014, currently postdoctoral fellow,  
NYU Medical Center  
Youngjoo Cho, Statistics, 2011 – 2014, currently assistant professor  
Division of Biostatistics, University of Wisconsin-Milwaukee School of Public Health  
Xiang Zhan, Statistics, 2011 – 2014, currently assistant professor  
Division of Biostatistics, Penn State College of Medicine

**University of Colorado:**

Manish Dalwani, Biostatistics and Informatics, 2014 – 2017, currently Co-Founder  
& Chief Scientific Officer at Vita Analytics Inc  
Kyle Smith, Computational Biosciences, 2016 – 2017, currently Bioinformatic Staff Scientist,  
St. Jude's Hospital, Memphis, Tennessee  
Alex Jensen, Biostatistics and Informatics, 2018 – present  
Kevin Josey, Biostatistics and Informatics, 2016 – 2020 (Joint with Fan Yang and Elizabeth Juaréz-  
Colunga)  
Charlie Carpenter, Biostatistics and Informatics, 2020 – present (Joint with Brandie Wagner)  
Ted Warsavage, Biostatistics and Informatics, 2020 – present  
Emily Mastej, Computational Biosciences, 2019 – present

**Post-Doctoral Students:****Penn State University:**

Yeonok Lee, Statistics, 2011 – 2014 (joint with Y. Zhang), currently statistical modeller, Develop-  
ment Office, St. Jude's Children Hospital.

**University of Colorado:**

Pratyaydipta Rudra, Biostatistics and Informatics (joint with K. Kechris), 2015 – 2018, currently  
Assistant Professor, Department of Statistics, Oklahoma State University  
Xuhong Zhang, Biostatistics and Informatics, 2017 – present (joint with Fuyong Xing)  
Efrén Cruz-Cortes, Biostatistics and Informatics, 2017 – 2019  
Olivier Simon, Biostatistics and Informatics, 2018 – present  
Tushar Ghosh, Biostatistics and Informatics, 2018 – present (joint with K. Kechris)  
Elin Shaddox, Biostatistics and Informatics, 2019 – present (joint with K. Kechris)  
Debmalya Nandy, Biostatistics and Informatics, 2019 – present (joint with K. Kechris)  
Zachary Richardson, Biostatistics and Informatics, 2019 – 2020  
Souvik Seal, Biostatistics and Informatics, 2020 – present (joint with K. Kechris)

Thao Vu, Biostatistics and Informatics, 2020 – present (joint with K. Kechris)

**Other Visitors:**

Hande Konsuk, Hacettepe University, 2012

**Thesis Committees:**

**University of Colorado:**

Elise Ameoka, Biostatistics and Informatics, 2017

Kayla Williamson, Biostatistics and Informatics, 2017

Cuining Liu, Biostatistics and Informatics, 2019

Emma Jones, Biostatistics and Informatics, 2019

**Dissertation Committees:**

**University of Michigan:**

Mike Epstein, Biostatistics, 2001 – 2002.

Nichole Carlson, Biostatistics, 2001– 2003.

Jasmanda Wu, Epidemiology, 2001– 2002.

Shibao Feng, Biostatistics, 2001 – 2003.

Lei Liu, Biostatistics, 2001 – 2004.

Stephanie Borchardt, Epidemiology, 2003 – 2004.

Debbie Lown, Epidemiology, 2002 – 2005.

Al Levin, Epidemiology, 2003 – 2005.

Adi Andrei, Biostatistics, 2003 – 2005.

Laura McEwen-Mattei, Epidemiology, 2003 – 2005.

Shona Dallal, Epidemiology, 2003 – 2005.

Abdullah AlSwuilem, Epidemiology, 2003 – 2005.

Chris Riolo, Epidemiology, 2000 – 2005.

Hsin-Jen Tsai, Nutritional Sciences, 2003 – 2006.

Patty Juliao, Epidemiology, 2002 – 2006.

Ying Kong, Epidemiology, 2004 – 2006.

Akarin Pharibul, Statistics, 2002 – 2006.

Dan Rhodes, Bioinformatics, 2004 – 2006.

Chad Creighton, Bioinformatics, 2004 – 2006.

Andrew Skol, Biostatistics, 2005 – 2006.

Rohit Kulkarni, Statistics, 2005 – 2006.

Aaron Sussell, Occupational and Environmental Health, 2002 – 2006.

Melissa Slotnick, Occupational and Environmental Health, 2004 – 2007.

Vivian Colon, Epidemiology, 2005 – 2007.

Kai Bullard, Epidemiology, 2005 – 2007.

Angela Liu, Biostatistics, 2006 – 2007.

**Penn State University:**

Sujana Ghosh, Biochemistry and Molecular Biology, 2008 – 2011

Ho Sung Rhee, Biochemistry and Molecular Biology, 2008 – 2011

Yan Zhuang, Agricultural Economics and Rural Sociology, 2009 – 2010

Jianping Sun, Statistics, 2009 – 2010  
Chris Groendyke, Statistics, 2009 – 2010  
Julia Jennings, Anthropology and Demography (Minor in Statistics), 2010  
Song Li, Integrative Biosciences, 2010  
Zhenhai Zhang, Bioinformatics, 2009 – 2011  
Kiranmoy Das, Statistics, 2010 – 2011  
Junyi Lin, Statistics, 2009 – 2011  
Duy Vu, Statistics, 2010 – 2013  
Qiuying Shen, Biology, 2011  
Jian Li, Biochemistry and Molecular Biology, 2009 – 2012  
Jialin Xu, Statistics, 2010 – 2012  
Qianyi Ma, Nutritional Sciences, 2010 – 2012  
Isaac Dialsingh, Statistics, 2010 – 2012  
Celine Han, Integrative Biosciences, 2011 – 2015  
Rene Koo Flasher, Accounting, 2011 – 2013  
Denise Finney, Crop and Soil Sciences, 2011 – 2015  
Anjel Helms, Ecology, 2011 – 2015  
Yao Zheng, Human Development and Family Studies, 2012 – 2014  
Neerja Katiyar, Integrative Biosciences, 2012 – 2013  
Vishesh Karwa, Statistics, 2012 – 2014  
Xizhen Cai, Statistics, 2012 – 2014

**University of Colorado:**

Bryan Vestal, Biostatistics and Informatics, 2015 – 2017  
Peter Dewitt, Biostatistics and Informatics, 2015 – 2017  
Junxiao Hu, Biostatistics and Informatics, 2016 – 2019  
KaraAnn Clouse, Health Services Research, 2014 – 2017  
Eline Van der Broek, Health Services Research, 2016 – 2017  
Angela Czaja, Pharmaceutical Sciences, 2017 – 2019  
Sarah Ryan, Biostatistics and Informatics, 2019 – 2020  
Yonghua Zhang, Biostatistics and Informatics, 2019 – present

**Lab Rotation Students:**

**University of Michigan:**

Paul Nicholas, Biostatistics, Winter 2003  
Kate Kirby, Biostatistics, Winter 2003  
Viktoriya Strumba, Bioinformatics, Winter 2004  
Omar Halawa, Bioinformatics, Winter 2006  
Bin Zhu, Biostatistics, Fall 2006  
Ali Shojaie, Statistics, Fall 2006, Winter 2007  
Matt Zawitowski, Biostatistics, Winter 2007

**Penn State University:**

Yang Liu, Bioinformatics, Fall 2010  
Scott Roths, Statistics, Summer 2010

Daisy Phillips, Statistics, Spring/Summer 2011  
Yihan Li, Statistics, 2011  
Yeying Zhu, Statistics, 2010–2011  
Bahaedinne Toufak, Statistics, 2012 – 2014

**University of Colorado:**

Mayla Boguslav, Computational Biosciences, 2017.  
Nicolle Witte, Computational Biosciences, 2017.  
Josephina Hendrix, Computational Biosciences, 2018.  
Stella Veazey, Biostatistics and Informatics, 2017– 2018.  
Emily Mastej, Computational Biosciences, 2019.  
Eric Prince, Computational Biosciences, 2021.  
Connor Ekhill, Computational Biosciences, 2021.  
Ksenia Lepikhina, Statistics (CU Boulder), 2021.  
Shuozhi Zuo, Biostatistics and Informatics, 2021 – present.

**Examination Committees:**

**University of Michigan:**

Xing Li, Bioinformatics, 2003  
Yili Chen, Bioinformatics, 2003  
Jianjun Yu, Bioinformatics, 2003 (chair)  
Carlos Santos, Bioinformatics, 2003  
Yu Chen, Bioinformatics, 2003  
Dan Rhodes, Bioinformatics, 2004 (chair)  
Pete Ulintz, Bioinformatics, 2004 (chair)  
Reagan Kelly, Bioinformatics, 2005  
Abhik Shah, Bioinformatics, 2005  
Barry Taylor, Bioinformatics, 2006 (chair)

**Penn State University:**

Akshay Katamansu, IBIOS, 2013  
Juan Antonio Raygoza Garay, IBIOS, 2013  
Yafei Lyu, IBIOS, 2014  
Rahulsimham Vegesna, IBIOS, 2014

**University of Colorado:**

Nicolle Witte, Computational Biosciences, 2017.  
Janet Siefert, Computational Biosciences, 2018.