

Debashis Ghosh

533 Fifth Street
Ann Arbor, MI 48103
(734) 997-7813

University of Michigan
Department of Biostatistics
1420 Washington Heights
Ann Arbor, MI 48109-2029
Phone: (734) 615-9824
FAX: (734) 763-2215
E-mail: ghoshd@umich.edu

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
		B.A., Statistics
		B.A., Economics
		B.A., French Studies

Academic Positions:

2005-present	Associate Professor, Department of Biostatistics, University of Michigan, Ann Arbor
2003-present	Assistant Member, UM Comprehensive Cancer Center
2002-present	Affiliate Faculty, Bioinformatics Program University of Michigan, Ann Arbor
2001-present	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

Awards:

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

Publications which have received peer review:

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.

Ghosh, D. and Lin, D.Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554–562.

Ghosh, D. (2000). Methods for the analysis of multiple events and death. *Controlled Clinical Trials* **21**, 115–126.

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.

Ghosh, D. (2001). Efficiency considerations in the additive hazards model with current status data. *Statistica Neerlandica* **55**, 367 – 376.

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.

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.

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.

Ghosh, D. and Chinnaiyan, A. M. (2002). Mixture modelling of gene expression data from microarray experiments. *Bioinformatics* **18**, 275–286.

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). α -Methylacyl Coenzyme A Racemase as a Tissue Biomarker for Prostate Cancer . *Journal of the American Medical Association* **287**, 1662–1670.
- Goldstein, D., Ghosh, D. and Conlon, E. (2002). Statistical issues in the clustering of gene expression data. *Statistica Sinica* **12**, 219–241.
- Ghosh, D. (2002). Resampling methods for variance estimation of singular value decomposition analyses from microarray experiments. *Functional and Integrative Genomics* **2**, 92 – 97.
- Ghosh, D. and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663 – 688.
- 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.
- 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.
- 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.
- 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.
- 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.
- Ghosh, D. (2003). Goodness of fit methods for the additive risk model in tumorigenicity experiments, *Biometrics* **59**, 721 – 726.
- 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.
- Smolkin, M. and Ghosh, D. (2003). Cluster stability scores for cancer subtypes in microarray experiments, *BMC Bioinformatics* **4**, 36 – 42.

- 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.
- Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events in the presence of dependent censoring, *Biometrics* **59**, 877 – 885.
- Ghosh, D. (2003). Penalized discriminant methods for the classification of tumors from microarray experiments, *Biometrics* **59**, 992 – 1000.
- 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-Methylacyl-CoA racemase and prostate cancer. *Journal of the National Cancer Institute* **96**, 834 - 843.
- 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.
- 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.
- 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. *Am J Med Genet B Neuropsychiatric Genetics* **127**, 85 – 89.
- 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.
- 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**, 9309-14.
- Ghosh, D. (2004). Mixture models for assessing differential expression in complex tissue using microarray data. *Bioinformatics* **20**, 1663 – 1669.
- Ghosh, D. (2004). Accelerated rates regression models for recurrent failure time data. *Lifetime Data Analysis* **10**, 247 – 261.
- Ghosh, D. (2004). Identification of shared components and sparse networks in gene expres-

sion data. *Journal of Virtual Learning Signal Integration and Signal Processing* **38**, 277 – 286.

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.

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.

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.

Ghosh, D. (2005). Nonparametric methods for identifying multiple replications of origins in genomewide data. *Functional and Integrative Genomics* **5**, 28 - 31.

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. 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.

Shen, R., Ghosh, D. and Chinnaiyan, A. M. (2005). Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. *BMC Genomics* **5**, 94.

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.

Abecasis, G. R., Ghosh, D. and Nichols, T.E. (2005). Linkage disequilibrium: ancient history drives the new genetics. *Human Heredity* **59**, 118-124.

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.

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.

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.

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.

Sen, S., Burmeister, M., and Ghosh, D. (2005). 5-HTTLPR and anxiety-related personality traits meta-analysis revisited: response to Munafo and colleagues. *Molecular Psychiatry* **10**, 893 – 895.

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.

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.

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.

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.

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.

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.

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.

Ghosh, D. (2006). Semiparametric inferences for the association parameter with semi-competing risks data, *Statistics in Medicine* **25**, 2059 – 2070.

Ghosh, D. (2006). Modelling tumor biology-progression relationships in screening trials,

Statistics in Medicine **25**, 1872 – 1884.

Ghosh, D. (2006). Random projection-based clustering for finding cancer subtypes in cancer microarray data, *Neurocomputing*, online January 19, 2006.

Elder, J. T., Ghosh, D. and Shah, R. B. (2006). Epidermal growth factor receptor (ErbB1) expression in prostate cancer progression: correlation with androgen independence. *The Prostate*, online June 1, 2006.

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.

Banerjee, M., Biswas, P., and Ghosh, D. (2006). Semiparametric binary regression under monotonicity constraints, *Scandinavian Journal of Statistics*, in press.

Ghosh, D. (2006). Semiparametric global cross-ratio models for bivariate censored data, *Scandinavian Journal of Statistics*, in press.

Ghosh, D. (2006). Shrunk p-values for assessing differential expression, with applications to genomic data analysis, *Biometrics*, in press.

MacDonald, J. W. and Ghosh, D. (2006). COPA - cancer outlier profile analysis, *Bioinformatics*, in press.

Ghosh, D. (2006). Incorporating monotonicity into the evaluation of a biomarker, *Biostatistics*, in press.

Slotnick, M. J., Meliker, J. R., Avruskin, G. A., Ghosh, D. and Nriagu, J. O. (2006). Toenails as a biomarker of inorganic arsenic intake from drinking water and foods. *Journal of Toxicology and Environmental Health*, in press.

Shen, R., Ghosh, D., Chinnaiyan, A. M. and Meng, Z. (2006). Eigengene-based linear discriminant analysis for gene expression data, *Bioinformatics*, in press.

Liu, D., Lin, X. and Ghosh, D. (2006). Semiparametric regression of multi-dimensional genetic pathway data: least squares kernel machines and linear mixed models, *Biometrics*, under revision.

Ghosh, D. and Chinnaiyan, A. M. (2006). Empirical Bayes identification of tumor progressor genes from microarray data, *Biometrical Journal*, under revision.

Chang, S., Ghosh, D., Linderman, J. L. and Kirschner, D. E. (2006). Length-based prediction of peptide-MHC class II binding affinity, *Bioinformatics*, under revision.

Choi, H., Shen, R., Chinnaiyan, A. M. and Ghosh, D. (2006). Mixture models for denoising of gene expression data across multiple microarray studies, *Statistical Applications in Genetics and Molecular Biology*, under revision.

Ghosh, D. (2006). Proportional hazards regression for cancer studies, *Biometrics*, under revision.

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*, under revision.

Ghosh, D., Rhodes, D. and Chinnaiyan, A. M. (2006). Combining genomic data in human studies, submitted.

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. (2006). Risk factors for mortality among patients with diabetes: the Translating Research into Action for Diabetes (TRIAD) study, submitted.

Shen, R., Ghosh, D. and Taylor, J. M. (2006). Analysis of tissue microarray data using measurement error models, submitted.

Choi, H. and Ghosh, D. (2006). A comparison of meta-analysis methods for gene expression data, submitted.

Ghosh, D. (2006). Semiparametric inference for surrogate endpoints with bivariate censored data, submitted.

Ghosh, D. (2006). Empirical Bayes methods for estimation and confidence intervals in high-dimensional problems, submitted.

Ghosh, D. (2006). Multiple testing procedures under confounding, submitted.

Yuan, Z. and Ghosh, D. (2006). Combining logistic regression models for multiple biomarkers, submitted.

In preparation

Sarkar, S., Zhou, T., and Ghosh, D. (2006). A general decision-theoretic approach to multiple testing procedures for false discovery and false nondiscovery rates, in preparation.

Poisson, L. M. and Ghosh, D. (2006). High-dimensional genomic data analysis, prognostic/predictive factors and levels of evidence, in preparation.

Poisson, L. M. and Ghosh, D. (2006). A statistical paradigm for integration of in vitro and

in vivo datasets for finding relevant gene expression profiles, in preparation.

Ghosh, D. (2006). Nonparametric and semiparametric inference for isotonic Poisson regression models, with application to spatial epidemiology, in preparation.

Chen, W., Ghosh, D., Raghunathan, T. E. and Sargent, D. (2006). A statistical model for individualized treatment based on genetic profiles.

Gu, W. J. and Ghosh, D. (2006). Global associations between copy number and transcript mRNA expression.

Gu, W. J., Shen, R., Choi, H. W., Chinnaiyan, A. M. and Ghosh, D. (2006). Mining genomic datasets in cancer studies.

Publications which have not received peer review:

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.

Chapters of books

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. (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. *Sta-*

tistical Methods in Medical Research **10**, 306–307.

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*, in press.

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

Current Grants

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

Statistical Methods for the Analysis of Functional Genomic Data

Total: \$150,000

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.

NIH/NHLBI P30 CA46592 (Wicha), 06/01/99 to 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.

American Cancer Society RSG-02-179-MGO (Chinnaiyan), 07/01/02 to 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 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.

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.

Completed Funding

MUNN IDEA Grant (Ghosh)

7/1/02 to 6/30/03

Total: \$20,000

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

Prostate SPORC Seed Grant (Ghosh)

7/1/02 to 6/30/03

Total: \$10,000

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

NIH/NCCR M01 RR00042 (Omenn), 03/01/01 to 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.

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.

Invited Talks

1. "Statistical Methods for Integration of Copy Number and transcript mRNA data":
 - Joint Statistical Meetings, Seattle, Washington, August 8, 2006.
2. "Statistical Methods for Analysis of Genomic/Proteomic Data in Complex Tissue":
 - Statistical Society of Canada, University of Western Ontario, London, Ontario, May 29, 2006.
3. "Meta-analysis of genomic data and multiple testing":
 - Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, March 9, 2006.
 - Department of Statistics, Pennsylvania State University, State College, PA, March 14, 2006.
4. "Introduction to Statistical Methods for Genomic Data Analysis":
 - Association of Molecular Pathology Annual Meeting, Phoenix, Arizona, November 11, 2005.

5. “Multiple testing and shrinkage estimation”:
 - Multiple Comparisons Procedures, Shanghai, China, August 18, 2005.
6. “Towards inference of chromosomal aberrations from genomic data”:
 - Joint Statistical Meetings, Minneapolis, Minnesota, August 10, 2005.
7. “Combining Genomic Data in Human Cancer Studies”:
 - Mathematical Biosciences Institute, Columbus, Ohio, April 22, 2005.
 - Division of Oncology Biostatistics, Department of Oncology, Johns Hopkins University, Baltimore, Maryland, March 1, 2006.
8. “Semiparametric Support Vector Machines for Gene Expression Data”:
 - Institute for Pure and Applied Mathematics, UCLA, June 4, 2004.
 - Joint Statistical Meetings, Toronto, Canada, August 8, 2004.
9. “Association Models for Bivariate Censored Data”:
 - Division of Biostatistics, Columbia University, March 4, 2004.
 - Division of Biostatistics, Cleveland Clinic, April 23, 2004.
10. “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.
11. “Statistical Methods for Clustering Microarray Data in Cancer Studies”:
 - First Canadian Workshop on Statistical Genomics, Toronto, Canada, September 3, 2003.
12. “Statistical Methods for the Analysis of Microarray Data”:
 - East Tennessee State University, January 22, 2004.
13. “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.

14. "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.
15. "Penalized Regression Models for the classification of tumors from microarray experiments":
 - Pacific Symposium on Biocomputing, Kauai, Hawaii, January 4, 2002.
 - Training Program in Bioinformatics, Texas A&M University, February 25, 2002.
 - Department of Biostatistics, University of Minnesota, May 18, 2002.
 - Institute for Pure and Applied Mathematics, UCLA, June 21, 2002.
 - Center for Molecular Medicine and Genetics, Wayne State University, February 20, 2003.
16. "Mixture modelling of microarray data," Statistics Department, University of Washington, July 13, 2001.
17. "Semiparametric Analysis of Recurrent Failure Time Data Using Accelerated Rates Models":
 - Statistics 2001 Canada Conference, Concordia University, Montreal, Canada, July 2, 2001.
 - Department of Epidemiology and Biostatistics, Imperial Cancer Research Fund, London, England, June 9, 2002.
18. "Semiparametric Analysis of Recurrent Failure Time Data and Dependent Censoring":
 - Fourth Triennial Symposium on Statistics, Calcutta University, Calcutta, India, December 27, 2000.
 - Joint Statistical Meetings, Atlanta, Georgia, August 7, 2001.
19. "Marginal Regression Models for Recurrent and Terminal Events":
 - Department of Statistics, University of Florida, Gainesville, Florida, February 17, 2000.
 - Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, February 24, 2000.
 - Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, March 3, 2000.

- Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, March 6, 2000.
 - Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, March 9, 2000.
 - Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, South Carolina, March 24, 2000.
 - Department of Biostatistics, M.D. Anderson Cancer Center, Houston, Texas, March 27, 2000.
20. “Nonparametric Analysis of Recurrent Events and Death,” Society of Clinical Trials Annual Meeting, Anaheim, California, May 3, 1999.
 21. “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.

Service

Teaching:

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

Biostatistics Department:

Faculty Search: 2000-2001, 2001-2002, 2005-2006
Candidacy: 2000-2001, 2001-2002, 2003 - 2004, 2005-2006 (chair)
Statistical Genetics Search Committee: 2001-2002, 2002-2003
Web: 2002-2003
Associate Director, Cancer Biostatistics Training Grant: 2005 – present

University of Michigan:

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 – present
Chair, Bioinformatics Pilot Grant Award Program: 2006
Genome Sciences Training Grant Faculty: 2001-present
Genome Sciences Training Grant Review Panel: 2002, 2004

External:

2000-2001: Referee for: *Biometrics*, *Statistics in Medicine*,
Journal of Multivariate Analysis, *Biostatistics*,
Statistics and Probability Letters, *Lifetime Data Analysis*

2001-2002: Referee for *Bioinformatics* (2 articles), *Biometrics*,
Biometrika, *Biostatistics*, *Genome Biology* (2 articles),
Lifetime Data Analysis and *JASA* (2 articles).

2002-2003: Referee for *Bayesian Statistics 7*, *Bioinformatics* (4 articles), *Biometrics*,
Biostatistics, *Journal of Computational and Graphical Statistics*,
BMC Bioinformatics (2 articles), *IMA Volume Series*, *Genome Biology*, *Nature Medicine*,
Diabetes Care, *Journal of Statistical Planning and Inference*, *Journal of Biological Chemistry*
JASA - CSA, *Comptes rendus biologiques*, *Neurocomputing* (2 articles).

2003-2004: Referee for *JASA - CSA*, *Bioinformatics* (3 articles), *Journal of Multivariate Analysis*,
American Journal of Pharmacogenomics, *Biostatistics*, *Computational Statistics and Data Analysis* (2 articles),
FEBS Letters, *Statistics in Medicine*, *Biometrics* (2 articles), book chapter reviewer, *Annals of Statistics*,
Aging Cell, *Comparative and Functional Genomics*

2004 - 2005: Referee for *Bioinformatics* (3 articles), *Biometrika*, *Biometrics*, *BMC Bioinformatics* (2 articles),
Statistical Applications in Genetics and Molecular Biology (2 articles), *Lifetime Data Analysis* (2 articles),
New England Journal of Medicine, *IEEE Transactions in Computational Biology and Bioinformatics*,
JASA - CSA (2 articles), *Journal of Clinical Oncology*, *Journal of Very Large Databases*, *Journal of Statistical Planning and Inference*

2005 - 2006: Referee for *Genome Biology* (2 articles), *JASA-CSA*, *Journal of Clinical Oncology*, *Journal of National Cancer Institute*, *BMC Bioinformatics*, *Biometrics*, *Applied Statistics*, *Lancet*, *Statistical Applications in Genetics and Molecular Biology* (2 articles), *Cancer Epidemiology, Biomarkers and Prevention*, *Journal of the Royal Statistical Society Series B*, *Bioinformatics* (3 articles), *JASA - Theory and Methods*

2006 - 2007: Referee for *PLOS Genetics*, *Nucleic Acids Research*, *American Journal of Management Science*, *Bioinformatics*, *Journal of Computational and Applied Mathematics*, *JASA-CSA*, *Journal of Clinical Oncology*, *Journal of Clinical Endocrinology and Metabolism*, *Statistical Advances in the Biomedical Sciences* (2 articles), *Meta-Analysis in Genetics* (book), *Annals of Statistics*

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.

NIH *ad hoc* Grant Reviewer, 2004, 2006.

NSA *ad hoc* Grant Reviewer, 2006.

University of Cyprus Grant Reviewer, 2006.

Reviewer, Springer-Verlag proposal, 2006.

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, ENAR, 2007.

Associate Editor, *Cancer Informatics*, 2005 - present

Associate Editor, *Journal of Bioinformatics - Theory and Application*, 2005 - present

Associate Editor, *BMC Bioinformatics*, 2005 - present

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

Associate Editor, *Biometrics*, 2006 - present

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 present)

Student Advising

Doctoral Students:

Primary Students:

Dawei Liu, Biostatistics, 2002 – 2005 (co-chair with X. Lin)
Wei Chen, Biostatistics, 2002 – 2006 (co-chair with T.E. Raghunathan)
Laila Poisson, Biostatistics, 2003 – present
Ronglai Shen, Biostatistics, 2002 – present (co-chair with J. Taylor)
Zheng Yuan, Biostatistics, 2004 – present
Hyungwon Choi, Biostatistics, 2006 – present
Jun Ding, Biostatistics, 2006 – present (co-chair with G. Abecasis)

Dissertation Committees:

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 AlSwailem, 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 – present
Melissa Slotnick, Occupational and Environmental Health, 2004 – present
Anne-Marie France, Epidemiology, 2005 – present
Vivian Colon, Epidemiology, 2005 – present
Kai Bullard, Epidemiology, 2005 – present
Carla Talarico, Epidemiology, 2005 – present
Yasemin Akdas, Epidemiology, 2006 – present
Jerry Jin, Environmental Health Sciences, 2006 – present
Julie Kim, Bioinformatics, 2006 – present.

Lab Rotation Students:

Viktoriya Strumba, Bioinformatics, Winter 2004

Omar Halawa, Bioinformatics, Winter 2006

Examination Committees:

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)