Curriculum Vitae of Wei Pan

August 2023

Address: A460 Mayo Building (MMC 303)
Minneapolis, MN 55455
PHONE: (612)626-2705, FAX: (612)626-0660
EMAIL: panxx014@umn.edu
WWW: `http://www.biostat.umn.edu/~weip`

Education: Ph.D. (Statistics), University of Wisconsin-Madison, June 1997.
Thesis title: "Non- and semi-parametric survival analysis with left truncated and interval censored data"
Thesis advisor: Professor Rick Chappell

M.S. (Computer Science), University of Wisconsin-Madison, May 1996.

M.S. (Statistics), University of Wisconsin-Madison, December 1995.

M.S.Eng. (Computer Engineering), Chinese Academy of Sciences, July 1991.

B.S.Eng. (Computer Science & Engineering) and B.S. (Applied Mathematics), Tsinghua University, July 1989. Double Degree Honor Program.

Experience: Professor, Division of Biostatistics
School of Public Health, University of Minnesota, July 2007–Present.
Interim Division Head, Aug 2017–July 2018.

Associate Professor, Division of Biostatistics
School of Public Health, University of Minnesota, June 2003–June 2007.

Assistant Professor, Division of Biostatistics
School of Public Health, University of Minnesota, July 1997–June 2003.
Associate member of the graduate faculty, September 1997–Present.

Faculty, Graduate Program in Bioinformatics, 2002–Present.

Member, Biomedical Genomics Center, 2002–Present.

Associate Member, Cancer Center, University of Minnesota, October 1997 –Present.

Research Assistant, Department of Biostatistics, University of Wisconsin

Medical School. January 1996 –June 1997.

Teaching Assistant, Department of Statistics, University of Wisconsin at Madison. September 1993–December 1995.

Software Engineer, The 6th Research Institute, Chinese Ministry of Mechanical and Electronic Industry. August 1991–May 1993.

Professional Activity

Study Section Member, NIGMS/NIH Biomedical Research and Research Training Review Subcommittee B (BRT-B), then renamed TWD-B, Sept 2011–June 2015.

Associate Editor, *Statistics in Biosciences*, Feb 2015–Present.

Associate Editor, *Journal of American Statistical Association*, August 2003–2006.

Editorial board of *Statistics in Medicine*, August 2002–2006.

Member, American Statistical Association (ASA), 1997–Present.

Member, Institute of Mathematical Statistics (IMS), 2009–Present.

Member, International Biometric Society (ENAR), July 1997–Present.

Member, International Genetic Epidemiology Society (IGES), 2015–Present.

Member, American Society of Human Genetics (ASHG), 2015–Present.

Honor and Award

1. Fellow, Institute of Mathematical Statistics (IMS), 2020.
2. Fellow, American Statistical Association (ASA), 2011.
3. Inductee to the Academy for Excellence in Health Research, "as the highest recognition of excellence in health science research by a full time faculty member at the University of Minnesota", 2020.
4. Elected member of Delta Omega Honorary Public Health Society, April 2001.
5. Pan, Xie and Shen (2010, *Biometrics* 66:474-484), 2010 *Biometrics* Best Paper Award, presented in the IBC 2012, Kobe, Japan, August 27, 2012.
6. Grant-in-Aid of Research, Artistry and Scholarship, University of Minnesota (July 1, 1998–Dec. 15, 1999).

Peer-reviewed Articles

1. Chen S, Lin Z, Shen X, Li L, **Pan W**. (2023). Inference of causal metabolite networks in the presence of invalid instrumental variables with GWAS summary data. *Genet Epidemiol*.

2. Ren J, Lin Z, **Pan W**. (2023). Integrating GWAS summary statistics, individual-level genotypic and omic data to enhance the performance for large-scale trait imputation. *Hum Mol Genet*, 32(17), 2693-2703.

3. Yao Y, Charkraborty D, Zhang L, Shen X; Alzheimer's Disease Neuroimaging Initiative; **Pan W**. (2023). Deep causal feature extraction and inference with neuroimaging genetic data. *Stat Med*, 42(20), 3665-3684.

4. Ren J, Lin Z, He R, Shen X, **Pan W**. (2023). Using GWAS summary data to impute traits for genotyped individuals. *HGG Adv*, 4(3), 100197.

5. Li C, Shen X, **Pan W**. (2023). Inference for a large directed acyclic graph with unspecified interventions. *JMLR*.

6. Chakraborty D, Zhuang Z, Xue H, Fiecas M, Shen X, **Pan W**. (2023). Deep learning-based feature extraction with MRI data in neuroimaging genetics for Alzheimer's disease. *Genes*, 14(3), 626. (A Special Issue on Molecular Biomarkers in Alzheimer's Disease).

7. Xue H, Shen X, **Pan W**. (2023). Causal inference in transcriptome-wide association studies with invalid instruments and GWAS summary data. *JASA*.

8. Lin Z, Xue H, **Pan W**. (2023). Robust multivariable Mendelian randomization based on constrained maximum likelihood. *American Journal of Human Genetics*.

9. Li C, Shen X, **Pan W**. (2023). Nonlinear causal discovery with confounders. *JASA*.

10. He R, Liu M, Lin Z, Zhuang Z, Shen X, **Pan W**. (2023). DeLIVR: a deep learning approach to IV regression for testing nonlinear causal effects in transcriptome-wide association studies. *Biostatistics*. 10.1093/biostatistics/kxac051.

11. Dai, B., Shen, X., Chen, L. Y., Li, C., **Pan, W**. (2023). Data-adaptive discriminative feature localization with statistically guaranteed interpretation. *Annals of Applied Statistics*.

12. Knutson KA, **Pan W**. (2023). MATS: A novel multi-ancestry transcriptome-wide association study to account for heterogeneity in the effects of cis-regulated gene expression on complex traits. *Hum Mol Genet.* doi: 10.1093/hmg/ddac247.

13. Spanbauer C, **Pan W**; ADNI, The Alzheimer's Disease Neuroimaging Initiative. (2023). Sparse prediction informed by genetic annotations using the logit normal prior for Bayesian regression tree ensembles. *Genet Epidemiol.* doi: 10.1002/gepi.22505.

14. Dai, B., Shen, X., **Pan, W**. (2022). Significance tests of feature relevance for a black-box learner. *IEEE Transactions on Neural Networks and Learning Systems*.

15. He R, Xue H, **Pan W**; Alzheimer's Disease Neuroimaging Initiative. (2022). Statistical power of transcriptome-wide association studies. *Genet Epidemiol*, 46(8), 572-588. doi: 10.1002/gepi.22491.

16. Lin Z, Knutson K, **Pan W**. (2022). Leveraging omic data to boost the power of genome-wide association studies. *Human Genetics and Genomics Advances*, 3. 100144. 10.1016/j.xhgg.2022.100144.

17. Liu M, Shen X, **Pan W**. (2022). Deep reinforcement learning for personalized treatment recommendation. *Stat Med*, 41(20), 4034-4056. doi: 10.1002/sim.9491.

18. Xue H, **Pan W** (2022) Robust inference of bi-directional causal relationships in presence of correlated pleiotropy with GWAS summary data. *PLoS Genet*, 18(5), e1010205.

19. Lin Z, Pan I, **Pan W**. (2022). A practical problem with Egger regression in Mendelian randomization. *PLOS Genetics*, 18, e1010166. 10.1371/journal.pgen.1010166.

20. Lin Z, Xue H, Malakhov MM, Knutson KA, **Pan W**. Accounting for non-linear effects of gene expression identifies additional associated genes in transcriptome-wide association studies. *Hum Mol Genet*. 2022 Jan 19:ddac015. doi: 10.1093/hmg/ddac015. Epub ahead of print. PMID: 35043938.

21. Lin Z, Deng Y, **Pan W**. (2021). Combining the strengths of inverse-variance weighting and Egger regression in Mendelian randomization using a mixture of regressions model. *PLoS Genetics*, 17 (11), e1009922. DOI: 10.1371/journal.pgen.1009922

22. Yang H, Zhuang Z, **Pan W**. (2021). A graph convolutional neural network for gene expression data analysis with multiple gene networks. *Statistics in Medicine*, 40 (25), 5547-5564. DOI: 10.1002/sim.9140

23. Wu, C., Zhu, J., King, A., Tong, X., Lu, Q., Park, J.Y., Wang, L., Gao, G., Deng, H.-W., Yang, Y., Knudsen, K.E., Rebbeck, T.R., Long, J., Zheng, W., **Pan, W.**, Conti, D.V., Haiman, C.A., Wu, L. (2021). Novel strategy for disease risk prediction incorporating predicted gene expression and DNA methylation data: a multiphased study of prostate cancer. *Cancer Communications*, 41 (12), 1387-1397. DOI: 10.1002/cac2.12205

24. Deng Y, **Pan W**. (2021). Model checking via testing for direct effects in Mendelian Randomization and transcriptome-wide association studies. *PLoS Computational Biology*, 17 (8), e1009266. DOI: 10.1371/journal.pcbi.1009266

25. Dai B, Shen X, **Pan W**. (2021). Two-level monotonic multistage recommender systems. *Electronic Journal of Statistics*, 15(2), 5545-5569. DOI: 10.1214/21-EJS1924

26. Xue H, Shen X, **Pan W**. (2021). Constrained Maximum Likelihood-Based Mendelian Randomization Robust to both Correlated and Uncorrelated Pleiotropic Effects. *American Journal of Human Genetics*, 108 (7), 1251-1269. DOI: 10.1016/j.ajhg.2021.05.014

27. Deng Y, He Y, Xu G, **Pan W** (in press). Speeding up Monte Carlo simulations for the adaptive sum of powered score test with importance sampling. To appear in *Biometrics*.

28. DiLernia A, Quevedo K, Camchong J, Lim K, **Pan W**, Zhang L (2022). Penalized model-based clustering of fMRI data *Biostatistics*, 23(3), 825-843.

29. Zhang L, DiLernia A, Quevedo K,Camchong J, Lim K, **Pan W.** (2021). A random covariance model for bi-level graphical modeling with application to resting-state fMRI data. *Biometrics*, 77(4), 1385-1396.

30. Liu M, Shen X, **Pan W**. (2021). Outcome weighted ψ-learning for individualizedtrea tment rules. *Stat*, 10:e343.

31. Deng Y, He Y, Xu G, **Pan W** (2020). Speeding Up Monte Carlo Simulations for the Adaptive Sum of Powered Score Test with Importance Sampling. To appear in *Biometrics*.

32. Xue H, **Pan W** (2020). Inferring Causal Direction Between Two Traits in the Presence of Horizontal Pleiotropy with GWAS Summary Data. *PLoS Genetics*, **16(11)**, e1009105.

33. Peng S, Shen X, **Pan W**. (2020). Reconstruction of a directed acyclic graph with intervention. *Electronic Journal of Statistics*, **14**, 4133-4164.

34. Knutson KA, Deng Y, **Pan W**; for the Alzheimer's Disease Neuroimaging Initiative (2020). Implicating Causal Brain Imaging Endophenotypes in Alzheimer's Disease using Multivariable IWAS and GWAS Summary Data. *NeuroImage*, **223**, 117347.

35. Yang T, Wei P, **Pan W** (2020). Integrative analysis of multi-omics data for discovering low-frequency variants associated with low-density lipoprotein cholesterol levels. *Bioinformatics*, **36**(21), 5223-5228.

36. Pattee J, **Pan W** (2020). Penalized regression and model selection methods for polygenic scores on summary statistics. *PLOS Computational Biology*, **16(10)**, e1008271.

37. Hebbel RP, Wei P, Milbauer L, Corban MT, Solovey A, Kiley J, Pattee J, Lerman LO, **Pan W**, Lerman A. (2020). Abnormal Endothelial Gene Expression Associated With Early Coronary Atherosclerosis. *Journal of the American Heart Association*, 9(14).

38. Wu C, Xu G, Shen X, **Pan W** (2020). A Regularization-Based Adaptive Test for High-Dimensional GLMs. *JMLR*, 21(129):1-42.

39. Xue H, Wu C, **Pan W** (2020). Leveraging existing GWAS summary data of genetically correlated and uncorrelated traits to improve power for a new GWAS. *Genet Epi*, **44**, 717-732.

40. Deng Y, **Pan W** (2020). A powerful and versatile colocalization test. *PLoS Comput Biol*, **16(4)**, e1007778.

41. Knutson KA, **Pan W** (in press). Integrating brain imaging endophenotypes with GWAS for Alzheimer's disease. To appear in *Quantitative Biology*.

42. He Y, Xu G, Wu C, **Pan W** (2020). Asymptotically Independent U-Statistics in High-Dimensional Testing. *Annals of Statistics*, **49**, 154-181.

43. Yang T, Wu C, Wei P, **Pan W** (2020). Integrating DNA sequencing and transcriptomic data for association analyses of low-frequency variants and lipid traits. *Human Molecular Genetics*, **29**, 515-526.

44. Wu C, **Pan W** (2020). A powerful fine-mapping method for transcriptome-wide association studies. *Human Genetics*, **139**, 199-213.

45. Xue H, **Pan W**, for the Alzheimer's Disease Neuroimaging Initiative (2020). Some statistical consideration in transcriptome-wide association studies. *Genetic Epidemiology*, **44**, 221-232.

46. Pattee J, Zhan X, Xiao G, **Pan W** (2020). Integrating germline and somatic genetics to identify genes associated with lung cancer. *Genetic Epidemiology*, **44**, 233-247.

47. Xiao M, Zhuang Z, **Pan W** (2020). Local Epigenomic Data are more Informative than Local Genome Sequence Data in Predicting Enhancer-Promoter Interactions Using Neural Networks. *Genes*, **11**, 41.

48. Yang T, Kim J, Wu C, Ma Y, Wei P, **Pan W** (2020). An adaptive test for meta-analysis of rare variant association studies. *Genet Epi*, **44**, 104-116.

49. Li C, Shen X, **Pan W** (2000). Likelihood Ratio Tests for a Large Directed Acyclic Graph. *JASA*, **115**, 1304-1319.

50. Wu C, **Pan W** (2019). Integration of methylation QTL and enhancer-target gene maps with schizophrenia GWAS summary results identifies novel genes. *Bioinformatics*, **35**, 3576-3583.

51. Zhuang Z, Shen X, **Pan W** (2019). A simple convolutional neural network for prediction of enhancer-promoter interactions with DNA sequence data. *Bioinformatics*, **35**, 2899-2906.

52. Xiao M, Shen X, **Pan W** (2019). Application of deep convolutional neural networks in classification of protein subcellular localization with microscopy images. *Genetic Epidemiology*, **43**,330-341.

53. Zhu Y, Shen X, **Pan W** (2020). On high-dimensional constrained maximum likelihood inference. *JASA*, **115** 217-230.

54. Yuan Y, Shen X, **Pan W**, Wang Z (2019). Constrained likelihood for reconstructing a directed acyclic Gaussian graph. *Biometrika*, **106**, 109-125.

55. Deng Y, **Pan W** (2018). Significance testing for allelic heterogeneity. *Genetics*, **210**, 25-32.

56. Austin E, **Pan W**, Shen X (2020). A new semiparametric approach to finite mixture of regressions using penalized regression via fusion. *Statistica Sinica*, **30**, 783-807.

57. Wu C, **Pan W** (2018). Integration of enhancer-promoter interactions with GWAS summary results identifies novel schizophrenia-associated genes and pathways. *Genetics*, **209**, 699-709.

58. Deng Y, **Pan W** (2018). Improved Use of Small Reference Panels for Conditional and Joint Analysis with GWAS Summary Statistics. *Genetics*, **209**, 401-408.

59. Wu C, Xu G. **Pan W**, for Alzheimer's Disease Neuroimaging Initiative (in press). An adaptive test on high-dimensional parameters in generalized linear models. To appear in *Statistica Sinica*.

60. Wu C, **Pan W** (2018). Integrating eQTL data with GWAS summary statistics in pathway-based analysis with application to schizophrenia. *Genetic Epidemiology*, **42**, 303-316.

61. Park JY, Wu C, Basu S, McGue M, **Pan W** (2018). Adaptive SNP-set association testing in generalized linear mixed models with application to family studies. *Behavior Genetics*, **48**, 55-66.

62. Deng Y, **Pan W** (2017). Testing genetic pleiotropy with GWAS summary statistics for marginal and conditional analyses. *Genetics*, **207**, 1285-1299.

63. Xu Z, Wu C, Wei P, **Pan W** (2017). A Powerful Framework for Integrating eQTL and GWAS Summary Data. *Genetics*, **207**, 893-902.

64. Xu Z, Wu C, **Pan W**; Alzheimer's Disease Neuroimaging Initiative (2017). Imaging-wide association study: Integrating imaging endophenotypes in GWAS. *Neuroimage*, **159**, 159-169.

65. Xu Z, Xu G, **Pan W**; Alzheimer's Disease Neuroimaging Initiative (2017). Adaptive testing for association between two random vectors in moderate to high dimensions. *Genet Epidemiol*, **41**, 599-609.

66. Liu B, Wu C, Shen X, **Pan W** (2017). A novel and efficient algorithm for de novo discovery of mutated driver pathways in cancer. *Ann Applied Statist*, **11**, 1481-1512.

67. Deng Y, **Pan W** (2017). Conditional analysis of multiple traits based on marginal GWAS summary statistics. *Genetic Epidemiology*, **41**, 427-436.

68. Gong W, Rasmussen T, Singh B, Koyano-Nakagawa N, **Pan W**, Garry D (2017). Dpath software reveals hierarchical hemato-endothelial lineages using of Etv2- progenitors based on single cell transcriptome analysis. *Nature Communications*, **8**, 14362.

69. Kim J, **Pan W**, and for the Alzheimer's Disease Neuroimaging Initiative (2017). Adaptive testing for multiple traits in a proportional odds model with applications to detect SNP-brain network associations. *Genetic Epidemiology*, **41**, 259-277.

70. Gao C, Kim J, **Pan W**, and for the Alzheimer's Disease Neuroimaging Initiative (2017). Adaptive testing of SNP-brain functional connectivity association via a modular network analysis. *Proceedings of the Pacific Symposium Biocomputing 2017*, 58-69.

71. Kwak I-Y, **Pan W** (2017). Gene- and pathway-based association tests for multiple traits with GWAS summary statistics. *Bioinformatics*, **33** 64-71.

72. Wu C, Kwon S, Shen X, **Pan W** (2016). A new algorithm and theory for penalized regression-based clustering. *Journal of Machine Learning Research*, **17(188)**, 1-25.

73. Xu G, Lin L, Wei P, **Pan W** (2016). An adaptive two-sample test for high-dimensional means. *Biometrika*, **103**, 609-624.

74. Wei P, Cao Y, Zhang Y, Xu Z, Kwak IY, Boerwinkle E, **Pan W** (2016). On robust association testing for quantitative traits and rare variants. *G3*, **6**, 3941-3950.

75. Kim J, Zhang Y, **Pan W**, for the Alzheimer's Disease Neuroimaging Initiative (2016). Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics*, **203**, 715-731.

76. Wu C, Chen J, Kim J, **Pan W** (2016). An adaptive association test for microbiome data. *Genome Medicine*, **8**, 56.

77. Gao C, Zhu Y, Shen X, **Pan W** (2016). Estimation of multiple networks in Gaussian mixture models. *Electronic Journal of Statistics*, **10**, 1133-1154.

78. Chen L, et al. (2016). Carotid intima-media thickness and arterial stiffness and the risk of atrial fibrillation: the Atherosclerosis Risk in Communities (ARIC) Study, Multi-Ethnic Study of Atherosclerosis (MESA), and the Rotterdam Study. *JAHA: Journal of the American Heart Association*, **5**, e002907.

79. Bekwelem W, et al. (2016). Carotid atherosclerosis and stroke in atrial fibrillation: the Atherosclerosis Risk in Communities (ARIC) Study. *Stroke*, **47**, 1643-1646.

80. Liu B, Shen X, **Pan W** (2016). Non-linear joint latent variable models and integrative tumor subtype discovery. *Statistical Analysis and Data Mining*, **9**, 106-116.

81. Xu Z, **Pan W** (2016). Binomial mixture model-based association testing to account for genetic heterogeneity for GWAS. *Genet Epidemiol.*, **40**, 202-209.

82. Liu B, Shen X, **Pan W** (2016). Integrative and regularized principal component analysis of multiple sources of data. *Statistics in Medicine*, **35**, 2235-2250.

83. Kwak I-Y, **Pan W** (2016). Adaptive gene- and pathway-trait association testing with GWAS summary statistics. *Bioinformatics*, **32**, 1178-1184.

84. Kim J, Bai Y, **Pan W** (2015). An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genet Epidemiol.*, **39**, 651-663.

85. Kim J, **Pan W**, for Alzheimer's Disease Neuroimaging Initiative (2015b). Highly adaptive tests for group differences in brain functional connectivity. *Neuroimage: Clinical*, **9**, 626-639.

86. Kim J, **Pan W**, for Alzheimer's Disease Neuroimaging Initiative (2015). A cautionary note on using secondary phenotypes in neuroimaging genetic studies. *Neuroimage*, **121**, 136-145.

87. **Pan W**, Kwak I-Y, Wei P (2015). A powerful pathway-based adaptive test for genetic association with common or rare variants. *American Journal of Human Genetics*, **97**, 86-98.

88. Xu Z, **Pan W**, for the Alzheimer's Disease Neuroimaging Initiative (2015). Approximate score-based testing with application to multivariate trait association analysis. *Genet Epidemiol.*, **39**, 469-479.

89. Austin E, Shen X, **Pan W** (2015). A novel statistic for global association testing based on penalized regression. *Genet Epidemiol.*, **39**, 415-426.

90. Kim J, Wozniak JR, Mueller BA, **Pan W** (2015). Testing group differences in brain functional connectivity: using correlations or partial correlations? *Brain Connect.*, **5**, 214-231.

91. **Pan W**, Chen YM, Wei P (2015). Testing for polygenic effects in genome-wide association studies. *Genet Epidemiol.*, **39**, 306-316.

92. Zhang Y, **Pan W** (2015). Principal component regression and linear mixed model in association analysis of structured sample s: competitors or complements? *Genet Epidemiol.*, **39**, 149-155.

93. Zhu Y, Shen X, **Pan W** (2014). Structural pursuit over multiple undirected graphs. *JASA*, **109**, 1683-1696.

94. Kim J, Wozniak JR, Mueller BA, Shen X, **Pan W** (2014). Comparison of statistical tests for group differences in brain functional networks. *NeuroImage*, **101**, 681-694.

95. **Pan W**, Kim J, Zhang Y, Shen X, Wei P (2014). A powerful and adaptive association test for rare variants. *Genetics*, **197**, 1081-1095.

96. Xu Z, Shen X, **Pan W**; Alzheimer's Disease Neuroimaging Initiative. (2014). Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS One*, **9**, e102312.

97. Kim S, **Pan W**, Shen X (2014). Penalized regression approaches to testing for quantitative trait-rare variant association. *Frontiers in Genetics*, **5**, 121. (Corresponding author; the first author was Dr Pan's student.)

98. Ho Y-Y, Baechler EC, Ortmann W, Behrens TW, Graham RR, Bhangale TR, **Pan W** (2014). Using gene expression to improve the power of genome-wide association analysis. *Human Heredity*, **78**, 94-103.

99. Zhang Y, Xu Z, Shen X, **Pan W**, for the Alzheimer's Disease Neuroimaging Initiative (2014). Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, **96**, 309-325. (Corresponding/senior author; the first two authors were Dr Pan's students.)

100. **Pan W**, Shen X, Liu B (2013). Cluster analysis: unsupervised learning via supervised learning with a non-convex penalty. *Journal of Machine Learning Research*, **14**, 1865-1889.

101. Zhang Y, Shen X, **Pan W** (2013). Adjusting for Population Stratification in a Fine Scale with Principal Components and Sequencing Data. *Genetic Epidemiology*, **37**, 787-801. (Corresponding/senior author; the first author was Dr Pan's student.)

102. Liu B, Shen X, **Pan W** (2013). Semi-supervised spectral clustering with application to detect population stratification. *Frontiers in Genetics*, 25. (Corresponding/senior author.)

103. Kim S, **Pan W**, Shen X (2013). Network-based penalized regression with application to genomic data. *Biometrics*, **69**, 582-593. (Corresponding/senior author; the first author was Dr Pan's student.)

104. Zhu Y, Shen X, **Pan W** (2013). Simultaneous grouping pursuit and feature selection over an undirected graph. *JASA*, **108**, 713-725.

105. Austin E, **Pan W**, Shen X (2013). Penalized regression and risk prediction in genome-wide association studies. *Statistical Analysis and Data Mining*, **6**, 315-328. (Corresponding/senior author; the first author is Dr Pan's student.)

106. Shen X, **Pan W**, Zhu Y, Zhou H (2013). On constrained and regularized high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, **65**, 807-832.

107. Yuan Y, Shen X, **Pan W** (2012). Maximum likelihood estimation over directed acyclic Gaussian graphs. *Statistical Analysis and Data Mining*, **5**, 523-530.

108. Austin E, **Pan W**, Shen X. (2014). Does the Inclusion of Rare Variants Improve Risk Prediction? *BMC Proceedings*, **8(Suppl 1)**, S94. (a special issue for the Genetic Analysis Workshop (GAW) 18.) (Senior author; the first author is Dr Pan's student.)

109. Zhang Y, **Pan W**. (2014). Adjusting for population stratification and relatedness with sequencing Data. *BMC Proceedings*, **8(Suppl 1)**, S42. (a special issue for the Genetic Analysis Workshop (GAW) 18.) (Senior author; the first author is Dr Pan's student.)

110. Zhang Y, Guan W, **Pan W** (2013). Adjustment for population stratification via principal components in association analysis of rare variants. *Genetic Epidemiology*, **37**, 99-109. (Corresponding/senior author; the first author is Dr Pan's student.)

111. Shen X, Huang H-C, **Pan W** (2012). Simultaneous supervised clustering and feature selection over a graph. *Biometrika*, **99**, 899-914.

112. Shen X, **Pan W**, Zhu Y (2012). Likelihood-based selection and sharp parameter estimation. *JASA*, **107**, 223-232.

113. Luo C, **Pan W**, Shen X (2012). A two-step penalized regression method with networked predictors. *Statistics in Biosciences* (a special issue on network data analysis), **4**, 27-46. (Corresponding/senior author; the first author was Dr Pan's student.)

114. Wei P, **Pan W** (2012). Bayesian joint modeling of multiple gene networks and diverse genomic data to identify target genes of a transcription factor. *Annals of Applied Statistics*, **6**, 334-355. (Senior author; the first author was Dr Pan's student.)

115. Oetting WS, Zhu Y, Brott MJ, Matas AJ, Cordner GK, **Pan W** (2012). Validation of genetic variants MTHFR, IL10 and TGFB1 with early acute rejection in kidney allograft transplantation. *Clinical Transplantation*, **26**, 418-423.

116. Han F, **Pan W** (2012). A composite likelihood approach to latent multivariate Gaussian modeling of SNP data with application to genetic association testing. *Biometrics*, **68**, 307-315. (Corresponding/senior author; the first author was Dr Pan's student.)

117. Montaniel KR, Billaud M, Graham C, Kim SK, Carlson M, Zeng W, Zeng O, **Pan W**, Isakson BE, Hall JL, Adhikari N (2012). Smooth muscle specific deletion of Ndst1 leads to decreased vessel luminal area and no change in blood pressure in conscious mice. *J Cardiovasc Transl Res.*, **5**, 274-279.

118. Wang H, Shen X, **Pan W** (2011). Large margin hierarchical classification with mutually exclusive class membership. *Journal of Machine Learning Research*, **12**, 2721-2748.

119. Basu S, **Pan W**, Shen X, Oetting B (2011) Multi-locus association testing with penalized regression. *Genetic Epidemiology*, **35**, 755-765. (Corresponding/senior author.)

120. **Pan W**, Basu S, Shen X (2011). Adaptive tests for detecting gene-gene and gene-environment interactions. *Human Heredity*, **72**, 98-109. (Corresponding/senior author.)

121. Basu S, **Pan W** (2011) Comparison of statistical tests for disease association with rare variants. *Genetic Epidemiology*, **35**, 606-619. (Corresponding/senior author.)

122. Basu S, **Pan W**, Oetting B (2011) A dimension reduction approach for modeling multi-locus interaction in case-control studies. *Human Heredity*, **71**, 234-245.

123. **Pan W**, Shen X (2011). Adaptive tests for association analysis of rare variants. *Genetic Epidemiology* **35**, 381-388. (Corresponding/senior author.)

124. **Pan W** (2011) Relationship between genomic distance-based regression and kernel machine regression for multi-marker association testing. *Genetic Epidemiology* **35**, 211-216.

125. Han F, **Pan W** (2011). Powerful multi-marker association tests: unifying genomic distance-based regression and logistic regression. *Genetic Epidemiology* **34**, 680-688. (Corresponding/senior author; the first author was Dr Pan's student.)

126. Wei P, Milbauer LC, Enenstein J, Nguyen J, **Pan W**, Hebbel RP. (2011). Differential endothelial cell gene expression by African Americans versus Caucasian Americans: a possible contribution to health disparity in vascular disease and cancer. *BMC Medicine* **9**:2.

127. Han F, **Pan W** (2010). A data-adaptive sum test for disease association with multiple common or rare variants. *Human Heredity* **70**, 42-54. (Corresponding/senior author; the first author was Dr Pan's student.)

128. Tai F, **Pan W**, Shen X (2010). Bayesian variable selection in regression with networked predictors. p.147-165, *Analysis of High Dimensional Data*, edited by T Cai and X Shen. Higher Education Press, Beijing, China. (Corresponding/senior author; the first author was Dr Pan's student.)

129. Zhou H, **Pan W**, Shen X (2009). Penalized model-based clustering with unconstrained covariance matrices. *Electronic Journal of Statistics* **3**, 1473-1496. (Corresponding/senior author; the first author was Dr Pan's student.)

130. Xie B, **Pan W**, Shen X (2010). Penalized mixtures of factor analyzers with application to clustering high dimensional microarray data. *Bioinformatics* **26**, 501-508. (Corresponding/senior author; the first author was Dr Pan's student.)

131. Xie Y, **Pan W**, Jeong KS, Xiao G, Khodursky KB (2010). A Bayesian approach to joint modeling of protein-DNA binding, gene expression and sequence data. *Statistics in Medicine* **29**, 489-503. (Corresponding/senior author; the first author was Dr Pan's student.)

132. **Pan W** (2010). Statistical tests of genetic association in the presence of gene-gene and gene-environment interactions. *Human Heredity* **69**, 131-142.

133. **Pan W**, Han F, Shen X (2010). Test selection with application to detecting disease association with multiple SNPs. *Human Heredity* **69**, 120-130.

134. **Pan W** (2010). A unified framework for detecting genetic association with multiple SNPs in a candidate gene or region: contrasting genotype scores and LD patterns between cases and controls. *Human Heredity* **69**, 1-13.

135. Wei P, **Pan W**. (2010). Network-based genomic discovery: application and comparison of Markov random field models. *Applied Statistics* **59**, 105-125. (Senior author; the first author was Dr Pan's student.)

136. Zhou H, **Pan W** (2009). Binomial mixture model-based association tests under genetic heterogeneity. *Annals of Human Genetics* **73**, 614-630. (Corresponding/senior author; the first author was Dr Pan's student.)

137. Zhu Y, **Pan W**, Shen X. (2009). Support vector machines with disease-gene-centric network penalty for high dimensional microarray data. *Statistics and Its Interface* **2**, 257-269. (Corresponding/senior author; the first author was Dr Pan's student.)

138. Wang J, Shen X, **Pan W**. (2009). On efficient large margin semisupervised learning: method and theory. *Journal of Machine Learning Research* **10**, 719-742.

139. **Pan W**. (2009). Network-based multiple locus linkage analysis of expression traits. *Bioinformatics* **25**, 1390-1396.

140. **Pan W**, Xie B, Shen X. (2010). Incorporating predictor network in penalized regression with application to microarray data. *Biometrics* **66**, 474-484.

141. Wang J, Shen X, **Pan W** (2009). On large margin hierarchical classification with multiple paths. *Journal of the American Statistical Association* **104**, 1213-1223.

142. **Pan, W.** (2009). Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genetic Epidemiology* **33**, 497-507.

143. Zhu, Y., Shen, X., **Pan, W.** (2009). Network-based support vector machine for classification of microarray samples. *BMC Bioinformatics* **10**(Suppl 1):S21. (Corresponding/senior author; the first author was Dr Pan's student.)

144. Parker, D.P., Brosseau, L., Samant, Y., **Pan, W.**, Xi, M., Haugan, D. and the Study Advisory Board (2009). A randomized-controlled intervention of machine guarding and related-safety programs in small metal-fabrication businesses. *Public Health Report* 124 Suppl 1, 90-100.

145. **Pan, W.** (2008). Network-based model weighting to detect multiple loci influencing complex diseases. *Human Genetics* **124**, 225-234.

146. Xie, B., **Pan, W.**, Shen, X. (2008). Variable selection in penalized model-based clustering via regularization on grouped parameters. *Biometrics* **64**, 921-930. (Corresponding/senior author; the first author was Dr Pan's student.)

147. Xie, B., **Pan, W.**, Shen, X. (2008). Penalized model-based clustering with cluster-specific diagonal covariance matrices and grouped variables. *Electronic Journal of Statistics* **2**, 168-212. (Corresponding/senior author; the first author was Dr Pan's student.)

148. Wei, P. and **Pan, W.** (2008). Incorporating gene networks into statistical tests for genomic data via a spatially correlated mixture model. *Bioinformatics* **24**, 404-411. (Corresponding/senior author; the first author was Dr Pan's student.)

149. **Pan, W.**, Wei, P. and Khodursky, A. (2008). A parametric joint model of DNA-protein binding, gene expression and DNA sequence data to detect target genes of a transcription factor. *Proceedings of the Pacific Symposium of Biocomputing* **13**, 465-476.

150. Milbauer, L.C., Wei, P., Enenstein, J., Hillery, C.A., Scott, J.P., Nelson, S.C., Jiang, A., Topper, J.N., Yang, R.B., Hirsch, B., **Pan, W.** and Hebbel, R.P. (2008). Genetic endothelial systems biology of sickle stroke risk. *Blood* **111**, 3872-3879.

151. Wei, P. and **Pan, W.** (2008). Incorporating gene functions into regression analysis of DNA-protein binding data and gene expression data to construct transcriptional networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **5**, 401-415. (Corresponding/senior author; the first author was Dr Pan's student.)

152. **Pan, W.**, Jeong, K.S., Xie, Y., Khodursky, A. (2008) A nonparametric empirical Bayes approach to joint modeling of multiple sources of genomic data. *Statistica Sinica* **18**, 709-729.

153. Tai, F. and **Pan, W.** (2007). Incorporating prior knowledge of gene functional groups into regularized discriminant analysis of microarray data. *Bioinformatics* **23**, 3170-3177.

154. Li, N., Wu, B., Wei, P., Xie, B., Xie, Y., Xiao, G. and **Pan, W.** (2007). Functional group-based linkage analysis of gene expression trait loci. *BMC Proceedings*, **1** (Suppl 1): S117.

155. Tai, F and **Pan, W.** (2007). Incorporating prior knowledge of predictors into penalized classifiers with multiple penalty terms. *Bioinformatics* **23**, 1775-1782.

156. Jiang, A., **Pan, W.**, Milbauer, L.C., Shyer, Y., Hebbel, R.P. (2007). A practical question based on cross-platform Microarray data normalization: are BOEC cells more like large vessel or microvascular endothelial cells or neither of them? *Journal of Computational Biology and Bioinformatics* **5**, 875-893.

157. **Pan, W.**, Shen, X. (2007). Penalized model-based clustering with application to variable selection. *Journal of Machine Learning Research* **22**, 1145-1164.

158. Xiao, G., **Pan, W.** (2007). Consensus clustering of gene expression data and its application to gene function prediction. *Journal of Computational and Graphical Statistics* **16**, 733-751.

159. **Pan, W.** and Shen, X. (2007). Semi-supervised learning via constraints. *Prediction and Discovery. Contemporary Mathematics vol. 443*, p.193-204. JS Verducci, X Shen and J Lafferty (editors). American Mathematical Society, Providence, Rhode Island.

160. Wang, J., Shen, X. and **Pan, W.** (2007). On transductive support vector machines. *Prediction and Discovery. Contemporary Mathematics vol. 443*, p.7-19. JS Verducci, X Shen and J Lafferty (editors). American Mathematical Society, Providence, Rhode Island.

161. Brosseau, L., Parker, D., Samant, Y., **Pan, W.** (2007). Mapping safety interventions in metalworking shops. *Journal of Occupational and Environmental Medicine* **49**, 338-345.

162. Xie, Y., **Pan, W.**, Jeong, K.S., Khodursky, A. (2007). Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data. *Statistics in Medicine* **26**, 2258-2275.

163. **Pan, W.**, Shen, X., Jiang, A. and Hebbel, R.P. (2006). Semi-supervised learning via penalized mixture model with application to microarray sample classification. *Bioinformatics* **22**, 2388-2395.

164. Xiao G, Martinez-Vaz BM, **Pan W**, Khodursky AB (2006). Operon information improves gene expression estimation for cDNA microarrays. *BMC Genomics* **7**:83.

165. Huang, D., **Pan, W.** (2006). Incorporating biological knowledge into distance-based clustering analysis of microarray gene expression data. *Bioinformatics* **22**:1259-1268.

166. Huang, D., Wei, P., **Pan, W.** (2006). Combining gene annotations and gene expression data in model-based clustering: a weighted method. *OMICS: A Journal of Integrative Biology* **10**:28-39.

167. **Pan, W.** (2006). Incorporating gene functions as priors in model-based clustering of microarray gene expression data. *Bioinformatics* **22**:795-801.

168. He, Y., **Pan, W.**, Lin, J. (2006). Cluster analysis using multivariate normal mixture models to detect differential gene expression with microarray data. *Computational Statistics & Data Analysis* **51**, 641-658.

169. **Pan, W.** (2006). Incorporating gene functional annotations in detecting differential gene expression. *Applied Statistics* **55**:301-316.

170. **Pan, W.**, Xiao G, Huang X (2006). Using input dependent weights for model combination and model selection with multiple sources of data. *Statistica Sinica* **16**:523-540.

171. DelGiudice, G.D., Fieberg, J., Riggs, M.R., Powell, M.C. and **Pan, W.** (2006) A long-term age-specific survival analysis of female white-tailed deer. *Journal of Wildlife Management* **70**, 1556-1568.

172. Samant Y, Parker D, Brosseau L, **Pan W**, Xi M, Haugan D. (2006). Profile of machine safety in small metal fabrication businesses. *American Journal of Industrial Medicine* **49**:352-359.

173. Liu J, Louis TA, **Pan W**, Ma JZ, Collins AJ. (2006). State-level adjusted ESRD incident rates: Use of observed vs model-predicted category-specific rates. *Kidney International* **69**:1459-1463.

174. Xie, Y., **Pan, W.**, Khodursky, A. (2005). A note on using permutation based false discovery rate estimate to compare different analysis methods for microarray data. *Bioinformatics* **21**:4280-4288.

175. Xiao, G. and **Pan, W.** (2005). Gene function prediction by a combined analysis of gene expression data and protein-protein interaction data. *Journal of Bioinformatics and Computational Biology* **3**:1371-1389.

176. Guo, X., **Pan, W.**, Connett, J.E., Hannan, P.J., French, S.A. (2005). Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in Medicine* **24**:3479-95.

177. Huang, X., **Pan, W.**, Grindle, S., Han, X., Chen, Y., Park, S.J., Miller, L.M., Hall, J. (2005). A comparative study of discriminating human heart failure etiology using gene expression profiles. *BMC Bioinformatics* **6**: 205.

178. Huang, X., **Pan, W.**, Han, X., Chen, Y., Miller, L.W., Hall, J. (2005). Borrowing information from relevant microarray studies for sample classification using weighted partial least squares. *Computational Biology and Chemistry* **29**, 204-211.

179. Martinez-Vaz, B.M., Xie, Y., **Pan, W.**, Khodursky, A.B. (2005) Genome-wide localization of mobile elements: experimental, statistical and biological considerations. *BMC Genomics* **6**: 81.

180. **Pan, W.** (2005). Incorporating biological information as a prior in an empirical bayes approach to analyzing microarray data. *Statistical Applications in Genetics and Molecular Biology* **4(1)**, Article 12.

181. Guo, X. and **Pan, W.** (2005). Using weighted permutation scores to detect differential gene expression with microarray data. *Journal of Bioinformatics and Computational Biology* **3**, 989-1006.

182. Luan, X., **Pan, W.**, Gerberich, S.G. and Carlin, B.P. (2005). Does it always help to adjust for misclassification of a binary outcome in logistic regression? *Statistics in Medicine* **24**, 2221-2234.

183. Xie, Y., Jeong, K.S., **Pan, W.**, Khodursky,A. and Carlin, B.P. (2004). A case study on choosing normalization methods and test statistics for microarray data. *Comparative and Functional Genomics* **5**, 432-444.

184. Munshi K, Parker D, Samant Y, Brosseau L, **Pan W**, Xi M. (2005). Machine safety evaluation in small metal working facilities: an evaluation of inter-rater reliability in the quantification of machine-related hazards. *American Journal of Industrial Medicine* **48**:381-388.

185. Huang, X., **Pan, W.**, Park, S., Han, X., Miller, L.W. and Hall, J. (2004). Modeling the relationship between LVAD support time and gene expression changes in the human heart by penalized partial least squares. *Bioinformatics* **20**, 888-894.

186. Liu, J., Louis, T.A., **Pan, W.**, Ma, J.Z. and Collins, A.J. (2004). Methods for estimating and interpreting provider-specific standardized mortality ratios. *Health Services and Outcomes Research Methodology* **4**, 135-149.

187. Huang, X. and **Pan, W.** (2003). Linear regression and two-class classification with gene expression data. *Bioinformatics* **19**, 2072-2078.

188. **Pan, W.**, Lin, J. and Le, C. (2003) A mixture model approach to detecting differentially expressed genes with microarray data. *Functional & Integrative Genomics* **3**, 117-124.

189. Guo, X., Qi, H., Verfaillie, C.M. and **Pan, W.** (2003). Statistical analysis of longitudinal gene expression data. *Bioinformatics* **19**, 1628-1635.

190. **Pan, W.** (2003). On the use of permutation in and the performance of a class of nonparametric methods to detect differential gene expression. *Bioinformatics* **19**, 1333-1340.

191. Qi, H., Aguiar, D.J., Williams, S.M., La Pean, A., **Pan, W.** and Verfaillie, C.M. (2003). Idenitification of genes responsible for osteroblast differentiation from human mesodermal progenitor cells. *Proceedings of National Academy of Science USA* **100**, 3305-3310.

192. Zhao, Y. and **Pan, W.** (2003). Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics* **19**, 1046-1054.

193. Lin, J., Ozeki, M., Javel, E., Zhao, Z.F., **Pan, W.**, Schlentz, E. and Levine, S. (2003). Identification of gene expression profiles in rat ears with cDNA microarrays. *Hearing Research* **175**, 2-13.

194. Lin, J., Tsuboi, Y., **Pan, W.**, Giebink, S.G. and Kim, Y. (2002). Analysis by cDNA microarrays of altered gene expression in middle ears of rats with pneumococcal infection. *International Journal Pediatric Otorhinolaryngology* **65**, 203-211.

195. **Pan, W.** (2002. A note on the use of marginal likelihood and conditional likelihood in analyzing clustered data. *The American Statistician* **56**, 171-174.

196. Lazovich, D., Parker, D.L., Brosseau, L.M., Milton, F.T., Dugan, S.K., **Pan, W.** and Hock, L. (2002) "Effectiveness of a Worksite Intervention to Reduce an Occupational Exposure: The Minnesota Wood Dust Study". *American Journal of Public Health* **92**, 1498-1505.

197. Huang, X. and **Pan, W.** (2002) Comparing three methods for variance estimation with duplicated high density oligonucleotide arrays. *Functional & Integrative Genomics* **2**, 126-133.

198. Murray, R.P., Connett, J.E., Rand, C.S., **Pan, W.** and Anthonisen, N.R. (2002). Persistence of the effect of the Lung Health Study (LHS) smoking intervention over eleven years. *Preventive Medicine* **35**, 314-319.

199. **Pan, W.**, Lin, J. and Le, C. (2002). How many replicates of arrays are required to detect gene expression changes in microarray experiments? a mixture model approach. *Genome Biology*, **3(5)**, research0022.1-10.

200. **Pan, W.** (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **12**, 546-554.

201. **Pan, W.**, Lin, J. and Le, C. (2002). Model-based cluster analysis of microarray gene expression data. *Genome Biology*, **3(2)**, research0009.1-8.

202. **Pan, W.** (2002). Approximate confidence intervals for one proportion and difference of two proportions. *Computational Statistics and Data Analysis* **40**, 143-157.

203. **Pan, W.** and Connett, J.E. (2002). Selecting the working correlation structure in generalized estimating equations with application to the Lung Health Study. *Statistica Sinica* **12**, 475-490.

204. **Pan, W.** and Chappell, R. (2002). Estimation in the Cox proportional hazards model with left truncated and interval censored data. *Biometrics* **58**, 64-70.

205. **Pan, W.** and Wall, M.M. (2002). Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Statistics in Medicine* **21**, 1429-1441.

206. **Pan, W.** (2002). Application of conditional moment tests to model checking for generalized linear models. *Biostatistics* **3**, 267-276.

207. **Pan, W.** (2002). Goodness-of-fit tests for GEE with correlated binary data. *Scandinavian Journal of Statistics* **29**, 101-110.

208. DelGiudice, G.D., Riggs, M.R., Joly, P. and **Pan, W.** (2002). Winter severity, survival and cause-specific mortality of female white-tailed deer in north central Minnesota. *The Journal of Wildlife Management* **66**, 698-717.

209. **Pan, W.** (2001). A multiple imputation approach to regression analysis for doubly censored data with application to AIDS studies. *Biometrics* **57**, 1245-1250.

210. **Pan, W.** (2001). On the robust variance estimator in generalised estimating equations. *Biometrika* **88**, 901-906.

211. **Pan, W.** (2001). Sample size and power calculations with correlated binary data. *Controlled Clinical Trials* **22**, 211-227.

212. **Pan, W.**, Connett, J.E., Porzio, G.C. and Weisberg, S. (2001). Graphical model checking with correlated response data. *Statistics in Medicine* **20**, 2935-2949.

213. **Pan, W.** (2001). Model selection in estimating equations. *Biometrics* **57**, 529-534.

214. **Pan, W.** (2001). Akaike's information criterion in generalized estimating equations". *Biometrics* **57**, 120-125.

215. **Pan, W.** (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis* **7**, 55-64.

216. **Pan, W.** and Connett, J.E. (2001). A multiple imputation approach to linear regression with clustered censored data. *Lifetime Data Analysis* **7**, 111-123.

217. **Pan, W.** and Le, C.T. (2001). Bootstrap model selection in generalized linear models. *Journal of Agricultural, Biological and Environmental Statistics* **6**, 49-61.

218. **Pan, W.** (2000). Smooth estimation of the survival for interval censored data. *Statistics in Medicine* **19**, 2611-2624.

219. **Pan, W.**, Connett, J.E. and Louis, T.A. (2000). A note on marginal linear regression with correlated response data. *The American Statistician* **54**, 191-195.

220. Brosseau, L.M., Parker, D., Lazovich, D., Dugan, S., Milton, T. and **Pan, W.** (2001). Inhalable dust exposures, tasks and use of ventilation in small woodworking shops: a pilot study. *American Industrial Hygiene Association Journal* **62**, 322-329.

221. **Pan, W.** (2000). A multiple imputation approach to Cox regression with interval censored data. *Biometrics* **56**, 192-203.

222. **Pan, W.** and Louis, T.A. (2000). A linear mixed-effects model for multivariate censored data. *Biometrics* **56**, 160-166.

223. **Pan, W.** (2000). A two-sample test with interval censored data via multiple imputation. *Statistics in Medicine* **19**, 1-11.

224. **Pan, W.** (1999). Shrinking classification trees for bootstrap aggregation. *Pattern Recognition Letters* **20**, 961-965.

225. **Pan, W.** and Kooperberg, C. (1999). Linear regression for bivariate censored data via multiple imputation. *Statistics in Medicine* **18**, 3111-3121.

226. **Pan, W.** (1999). Bootstrapping likelihood for model selection with small samples. *Journal of Computational and Graphical Statistics* **8**, 687-698.

227. **Pan, W.** and Chappell, R. (1999). A note on inconsistency of NPMLE of the distribution function from left-truncated and interval-censored data. *Lifetime Data Analysis* **5**, 281-291.

228. **Pan, W.** and Louis, T.A. (1999). "Two semi-parametric empirical Bayes estimators. *Computational Statistics and Data Analysis* **30**, 185-196.

229. **Pan, W.** (1999). A comparison of some two-sample tests with interval censored data. *Journal of Nonparametric Statistics* **12**, 133-146.

230. **Pan, W.** (1999). Extending the iterative convex minorant algorithm to the Cox model for interval censored data. *Journal of Computational and Graphical Statistics* **8**, 191-200.

231. **Pan, W.** (1998). Rank invariant tests with left truncated and interval censored data. *Journal of Statistical Computation and Simulation* **61**, 163-174.

232. **Pan, W.** and Chappell, R. (1998). A nonparametric estimator of survival functions for arbitrarily truncated and censored data. *Lifetime Data Analysis* **4**, 187-202.

233. **Pan, W.** and Chappell, R. (1998). Computation of the NPMLE of distribution functions for interval censored and truncated data with applications to the Cox model. *Computational Statistics and Data Analysis* **28**, 33-50.

234. **Pan, W.** and Chappell, R. (1998). Estimating survival curves with left-truncated and interval-censored data under monotone hazards. *Biometrics* **54**, 1053-1060.

235. **Pan, W.** and Chappell, R. (1998). Estimating survival curves with left-truncated and interval-censored data via the EMS algorithm. *Communications in Statistics – Theory and Methods* **27**, 777-793.

236. **Pan, W.**, Chappell, R. and Kosorok, M.R. (1998). On consistency of the monotone MLE of survival for left truncated and interval censored data. *Statistics and Probability Letters* **38**, 49-57.

237. **Pan, W.**, Li, L. and Zhang, X. (1991). Words segmentation in Chinese handwritten documents. *Proc. of the Chinese National Conference on Artificial Intelligence and Pattern Recognition.* 300-303, Harbin, China. (In Chinese)

238. Li, L., **Pan, W.** and Zhang, X. (1991). An algorithm of segmenting characters in scripted documents. *Proc. of the Chinese National Conference on Artificial Intelligence and Pattern Recognition.* 304-307, Harbin, China. (In Chinese)

239. **Pan, W.**, Jiang, X., Li, L. and Zhang, X. (1991). K×K thinning. *Proc. of the Second Chinese National Conference on Computer Vision and Intelligent Control.* 248-252, Wuhan, China. (In Chinese)

240. **Pan, W.**, Jiang, X., Zhang, X. (1990). Preprocessing of handwritten Chinese character recognition. *Proc. of Computer Science Graduates Conference*, Chinese Academy of Sciences, Beijing, China. (In Chinese)

Books and chapters in books

1. The United States Renal Data System, *USRDS 2002 Annual Data Report*, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, May 2002. (As a co-investigator in the USRDS Coordinating Center)

2. The United States Renal Data System, *USRDS 2001 Annual Data Report*, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, May 2001. (As a co-investigator in the USRDS Coordinating Center)

3. The United States Renal Data System, *USRDS 2000 Annual Data Report*, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, June 2000. (As a co-investigator in the USRDS Coordinating Center)

Non-peer reviewed Articles

1. **Pan, W.** (1998) "Bias/Variance Tradeoff in Combining Subsample Estimates for a Very Large Data Set". *Computing Science and Statistics* **30**, 379-381.

2. **Pan, W.** (1998) "Bagging Empirical Bayes Estimators". *Proceedings of the Section on Bayesian Statistical Science of the American Statistical Association*, 28-31.

Presentations and Abstracts

Invited Presentations:

1. Joint Conference on Statistics and Data Science (JCSDS), Beijing, China, July 11, 2023
2. ICSA, Chengdu, China, July 1, 2023
3. ENAR, Nashville, TN, March 21, 2023.
4. Poster highlight, IGES 2022 Annual Meeting, Paris, France, Sept 8, 2022.
5. BIRS-UBC-O Workshop on Deep Learning for Genetics, Genomics and Metagenomics: Latest developments and New Directions, Virtual, June 9, 2022.
6. Workshop on Statistical Methods in Genetic/Genomic Studies, IMS, National University of Singapore, Virtual, Jan 6, 2022.
7. Quantitative Biology, Virtual, Nov 12, 2021.
8. Dept of Statistics, Oregon State U, Virtual, Oct 25, 2021.
9. JSM, virtual meeting, August 12, 2021.
10. WNAR, virtual meeting, June 14, 2021.
11. ENAR, virtual meeting, March 2021.
12. JSM, virtual meeting, August 2020.
13. ENAR, virtual meeting, March 2020.
14. JSM, Denver, CO, August 2019.
15. Biostatistics Workshop, Georgia State University, Atlanta, GA, May 2019.
16. CMStatistics, London, UK, Dec 2019.
17. CMStatistics, Pisa, Italy, Dec 2018.
18. Dept of Biostatistics, U of Michigan, Dec 2018.
19. Dept of Statistics, NCSU, Nov 2018.
20. IGES, San Diego, CA, Oct 2018.
21. JSM, Vancouver, BC, Canada, July 2018.
22. Sichuan University, Chengdu, China, July 2018.

23. Southwestern Financial and Economic University, Chengdu, China, July 2018.
24. ICSA China Conference, Qingdao, China, July 2018.
25. Shanghai Jiaotong University, Shanghai, China, June 2018.
26. ENAR, Atlanda, GA, March 2018.
27. Dept of Computer Science, U of Minnesota, Nov 2017.
28. ASHG, Orlando, FL, Oct 2017.
29. ICSA Applied Symposium, Chicago, IL, June 2017.
30. ENAR, Washington DC, March 2017.
31. GAW 20, San Diego, CA, March 2017.
32. Dahshu 2017 Data Science & Computational Precision Health, San Francisco, CA, Feb 2017.
33. Institute of Bioinformatics, University of Georgia, GA, Nov 4, 2016.
34. ICSA Applied Statistics Symposium, Atlanda, GA, June 2016.
35. ENAR, Austin, TX, March 2016.
36. Mathematical and Statistical Challenges in Neuroimaging Data Analysis Workshop, Banff, Canada, Jan 31-Feb 5.
37. Center of Computational Biology and Bioinformatics, Indiana University, Dec 7, 2015.
38. JSM, Seattle, WA, August 2015.
39. Biostatistics and Bioinformatics Branch, NICHD, NIH, April 16, 2015.
40. Big Data Workshop, Shanghai, China, Nov 2014.
41. Department of Biotatistics, University of North Carolina at Chapel Hill, Sept 25, 2014.
42. JSM, Boston, MA, August 2014.
43. IMS, Sydney, Australia, July2014 2014.
44. Department of Biostatistics, U of Washington, April 24, 2014.
45. Division of Biostatistics, U of Texas Health Science Center at Houston, March 28, 2014.
46. ENAR, Baltimore, MD, March 19, 2014.
47. Department of Statistics, Michigan State U, Oct 29, 2013.
48. Department of Biotatistics, U of Iowa, Sept 30, 2013.
49. JSM, Montreal, Canada, August 7, 2013.
50. IMS-China Meeting, Chengdu, China, July 1, 2013.
51. Department of Statistics, Penn State U, April 18, 2013.
52. Department of Biostatistics, U of Wisconsin-Madison, April 5, 2013.

53. Biostatistics, NYU, March 18, 2013.
54. "A network-based penalized regression method with application to genomic data". Biostatistics, UPenn, Sept 18, 2012.
55. "Incorporating Predictor Network in Penalized Regression with Application to Microarray Data". *Biometrics* Best Paper Award Presentation, IBC 2012, Kobe, Japan, August 27, 2012.
56. "Adjustment for Population Stratification via Principal Components in Association Analysis of Rare Variants". Joint Statistical Meetings, San Diego, CA, July 27, 2012.
57. "Adjustment for Population Stratification via Principal Components in Association Analysis of Rare Variants". The 2nd Biostatistics Symposium, Beijing, China, July 8, 2012.
58. "Adjustment for Population Stratification via Principal Components in Association Analysis of Rare Variants". ICSA Applied Statistics Symposium, Boston, MA, June 24, 2012.
59. "Testing for Disease-Rare Variant Association with Sequence Data". Dept of Statistics, Purdue University, October 28, 2011.
60. "Testing for Disease-Rare Variant Association with Sequence Data". Institute of Health Informatics, University of Minnesota, October 6, 2011.
61. "Multi-locus Association Testing: Global Tests vs Penalized Regression". JSM, Miami Beach, August 2011.
62. "Multi-locus Association Testing: Global Tests vs Penalized Regression". ICSA Symposium of Applied Statisics, New York City, June 2011.
63. Why to "ignore" correlations: applications to genetic association analysis. Dept of Biostatistics, MD Anderson Cancer Center, March 2, 2011.
64. Why to "ignore" correlations: applications to genetic association analysis. Dept of Statistics, Yale University, Jan 24, 2011.
65. The Eighth ICSA International Conference, Guangzhou, China, December 19-22, 2010.
66. Joint Statistical Meeting, Vancouver, Canada, July 31-Aug 5, 2010.
67. First Joint Biostatistics Symposium, Beijing, China, July 15-18, 2010.
68. "Network-based Penalized Regression with Application to eQTL Analysis". International Biometric Society-ENAR, New Orleans, LA, March 21-23, 2010.
69. "Statistical Tests of Genetic Association with Multiple SNPs". Dept of Statistics, Columbia University, Nov 9, 2009.
70. "Statistical Tests of Genetic Association with Multiple SNPs". Division of Biostatistics, Albert Einstein College of Medicine, Oct 1, 2009.
71. "Bayesian Variable Selection in Regression with Networked Predictors". International Conference on Financial Statistics and Financial Econometrics–Nonparametric Modelling and Its Application, Chengdu, China, July 8-10, 2009.

72. "Bayesian Variable Selection in Regression with Networked Predictors". Workshop on CTW: Systems Biology of Biological Processes and Diseases: Biological Problems and Statistical Solutions, Mathematical Biosciences Institute, Ohio State University, April 17, 2009.

73. "Network-based Penalized Regression with Application to eQTL Analysis". Dept of Mathematics and Statistics, U of Minnesota-Duluth, April 10, 2009.

74. "Network-based Penalized Regression with Application to eQTL Analysis". Dept of Molecular and Computational Biology, University of Southern California, April 1, 2009.

75. "Some Old and New Tests of Disease Association with Multiple SNPs in Linkage Disequilibrium". Division of Biostatistics, School of Public Health, University of Minnesota, Dec 3, 2008.

76. "Penalized model-based clustering with application to microarray data". Joint Statistical Meetings, August 2-8, 2008.

77. "Network-based Penalized Regression with Application to eQTL Analysis". International Conference Machine Learning and Data Mining, Beijing, China, June 14–19, 2008.

78. "Statistical genomics and spatial statistics: Incorporating biological knowledge of genes into analysis of genomic data". Department of Statistics, Iowa State University, April 28, 2008.

79. "A parametric joint model of DNA-protein binding, gene expression and DNA sequence data to detect target genes of a transcription factor". Pacific Symposium of Biocomputing (PSB), Big Island, Hawaii, Jan 4-8, 2008.

80. "Incorporating biological knowledge of genes into statistical analysis of genomic data". Keynote address, Critical Assessment of Microarray Data Analysis (CAMDA), Valencia, Spain, December 12-14, 2007.

81. "Variable selection in penalized model-based clustering via regularization on grouped parameters". 2007 Nonparametric Statistics Conference *"Current and Future Trends in Nonparametrics*, University of South Carolina, Columbia, SC, October 10-12, 2007.

82. "A nonparametric empirical Bayes approach to joint modeling of DNA-protein binding data and gene expression data". NSF Sponsored International Conference on Bioinformatics, Hangzhou, China, June 11-14, 2007.

83. "Unsupervised and semi-supervised learning for high-dimensional data". International Workshop on Scientific Computing: Models, Algorithms and Applications, Hong Kong, 7-9 December 2006.

84. "Semi-supervised learning via penalized mixture model". St Jude Children's Research Hospital, Memphis, TN, Nov 2006.

85. "Input-dependent weight for model selection and model combination". Joint Statistical Meeting (JSM), Seattle, WA, August 2006.

86. "Semi-supervised learning via penalized mixture model". 2006 AMS-IMS-SIAM Summer research Conference on Machine and Statistical Learning, Snowbird Resort, Utah, June 2006.

87. "Joint modeling of gene expression data and DNA-protein binding data". Department of Statistics, Texas A&M University, March 2006.

88. "Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data". International Chinese Statistical Association (ICSA) Symposium, Washington DC, June 2005.

89. "Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data". NHLBI Biostatistics, Washington DC, June 2005.

90. "Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data". Division of Biostatistics, Columbia University, March 2005.

91. "Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data". Department of Biostatistics, University of Pittsburg, March 2005.

92. "Joint Modeling of DNA-Protein Binding Data and Gene Expression Data". Division of Biostatistics, Brown University, Feb 2005.

93. "Joint Modeling of DNA-Protein Binding Data and Gene Expression Data". Department of Biostatistics, University of Michigan, Feb 2005.

94. "Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data". Division of Biostatistics, Yale University, Nov 2004.

95. "Incorporating prior information via shrinkage: a combined analysis of genome-wide location data and gene expression data". Society of Chinese Bioscientists in America (SCBA) Conference, Beijing, China, July 2004.

96. "Statistical significance analysis of longitudinal gene expression data". International Chinese Statistical Association (ICSA) Symposium, San Diego, CA, June 2004.

97. "Consensus clustering for gene function prediction", Interface, Baltimore, May 2004.

98. "Statistical methods for sample classification and prediction with microarray gene expression data", Department of Statistics, University Of Georgia, March 2004.

99. "Statistical methods for sample classification and
prediction with microarray gene expression data". Departments of Biostatistics, John Hopkins University, December 2003.

100. "Tumor classification with microarray data". 2003 NSF-RCN Retreat, September 2003, Mohonk Mountain House, NY.

101. "A weighted permutation method in detecting differential gene expression". RSS Workshop on Statistical Analysis of Gene Expression Data, July 11-14 2003, Wye College Conference Center, Kent, UK.

102. "Statistical analysis of longitudinal gene expression data". Craybill Bioinformatics Conference at Colorado State University, June 2003.

103. "Analysis of gene expression data". the ASA Local Chapter Meeting, Minneapolis, December 2002.

104. "On clustered data analysis". School of Public Health, Ohio State University, November 2002.

105. "Longitudinal gene expression data analysis". Bioinformatics Center, Ohio State University, November 2002.

106. "Detecting differential gene expression". Health Informatics Seminar, co-sponsored by the BioTechnology Institute and the Mathematics Colloquium, University of Minnesota, September 2002.

107. "Sample size calculations for microarray data". 2002 Conference of Biomedical Genomics Center, University of Minnesota June 2002.

108. "Statistical Methods for Discovering Differentially Expressed Genes in Replicated Microarray Experiments". Departments of Biochemistry and Applied Mathematics & Statistics, University of Minnesota-Duluth, November 2001.

109. "Statistical Methods for Discovering Differentially Expressed Genes in Replicated Microarray Experiments". Division of Biostatistics, University of Minnesota, September 2001.

110. "A Multiple Imputation Approach to Regression Analysis for Doubly Censored Data with Application to AIDS Studies". Statistics Department, University of Missouri, March 2001.

111. "Inconsistency of Survival Functions Estimated from The Proportional Hazards Model with Left Truncated and Interval Censored Data". A Special Contributed Session, International Biometric Society ENAR Annual Meeting, March 2000, Chicago, IL.

112. "A Two-Sample Test with Interval Censored Data via Multiple Imputation". Upper-Midwest Biostatistics Symposium, August 23, 1999, Rochester, MN.

113. "Estimation in the Cox Proportional Hazards Model with Left Truncated and Interval Censored Data". University of Minnesota, School of Statistics, December 3, 1998.

114. "Selecting the Working Correlation Structure in Generalized Estimating Equations with Application to the Lung Health Study". University of Minnesota, Division of Biostatistics, School of Public Health, November 17, 1998.

115. Rensselear Polytechnic Institute, Department of Decision Sciences and Engineering Systems, February 1997.

116. University of Minnesota, Division of Biostatistics, February 1997.

117. State University of New York–Stony Brook, Department of Applied Mathematics and Statistics. February 1997.

Contributed Conference Presentations:

1. "Selecting the Working Correlation Structure in Generalized Estimating Equations with Application to Lung Health Study". Joint Statistical Meeting, August 1999, Baltimore, MD.
2. "Selecting Working Correlation Structure in Generalized Estimating Equations with Application to Lung Health Study". Joint Statistical Meeting, August 1999, Baltimore, MD.
3. "A Multiple Imputation Approach to Linear Regression with Clustered Censored Data". International Biometric Society ENAR Annual Meeting, March 1999, Atlanta, GA.
4. "Bagging Empirical Bayes Estimators". Joint Statistical Meeting, August 1998, Dallas, TX.
5. "Estimation in the Cox Proportional Hazards Model with Left Truncated and Interval Censored Data". International Biometric Society WNAR Annual Meeting, June 1998. San Diego, CA.
6. "Bias/Variance Tradeoff in Combining Subsample Estimates for a Very Large Data Set", Statistics and Computing Science Interface'98, May 1998. Minneapolis, MN.
7. "Nonparametric Estimation of the Survival with Left-truncated and Interval-censored Data", International Biometric Society ENAR Annual Meeting, March 1997. Memphis, TN.

Conference Abstracts:

1. Hebbel RP, Jiang AX, Hillery CA, Scott JP, Nelson SC, Enenstein J, Milbauer LM, Bodempudi V, **Pan W**, Topper JN, Yang RB. (2005). "Genetic influence on the systems biology of sickle stroke risk detected by endothelial gene expression". *Blood* **106**:26A-26A.
2. Oetting WS, Brott MJ, **Pan W**, Kandaswamy R, Matas AJ. (2005). "The -238A allele of TNF alpha is associated with acute rejection in kidney transplantation". *American Journal of Transplantation* **5**:303-303 Suppl.
3. Xue, J.L., Ebben, J., Snyder, J., Chen, S.C., Ma, J., **Pan, W.**, Collins, A.J. (2000). "Geographic and clinical comparison between peritoneal and hemodialysis patient". *American Journal of Kidney Diseases* **35**, A29.
4. Xue, J.L., Snyder, J., Ebben, J., Chen, S.C., Ma , J., **Pan, W.**, Collins, A.J. (2000). "Peritoneal dialysis versus hemodialysis: Differences in transplantation and modality change". *American Journal of Kidney Diseases* **35**, A29.

PI/co-PI on External Grants

1. MPI, (MPIs: X Shen, W Pan), Estimation and inference in directed acyclic graphical models for biological networks, NIH R01 AG074858, 4/01/22 - 01/31/27.
2. PI, Causal and integrative deep learning for Alzheimer's disease genetics, NIH U01 AG073079, 9/15/21-8/31/26.
3. (Contact) PI, Deep Learning with Neuroimaging Genetic Data for Alzheimer's disease, NIH R01 AG069895 (MPIs: W Pan, X Shen), 09/30/20-6/30/25.
4. PI, Integrating Alzheimer's disease GWAS with proteomic and metabolomic QTL data, NIH RF1 AG067924, 9/15/20-8/31/24.
5. PI, Discovering causal genes, brain regions and other risk factors for Alzheimer's disease, NIH R01 AG065636, 3/1/20-2/21/25.
6. PI, Biostatistics in Genetics and Genomics Training Program, T32GM132063 (MPIs: W Pan, S Basu), 07/20-06/25.
7. PI, Biostatistics in Genetics and Genomics Training Program, T32GM108557, 09/14-08/19, NCE to 7/20.
8. PI, Collaborative Research: Adaptive testing and rare-events analysis of high dimensional data, NSF DMS 1711226, 8/15/17-7/31/20, NCE to 7/21.
9. Co-PI, (MPIs: X Shen, W Pan), Estimation and Inference of Gene Regulatory Networks, NIH R01 GM126002, 09/15/17-08/30/21, NCE to 8/24.
10. Contact PI, (MPIs: W Pan, P Wei), Association Analysis of Rare Variants with Sequencing Data, NIH RO1 HL116720, 09/13-07/21, NCE to 7/24.
11. Contact PI, (MPIs: W Pan, X Shen), Genetic Association and Personalized Medicine, NIH RO1 HL105397, 03/11-04/19, NCE to 04/21.
12. PI, Statistical Methods for Genomic Data, NIH RO1 HL065462, 06/05–05/14. NIH R01 GM113250, 09/14–08/18, NCE to 08/21.
13. PI, Integrating Genomic and Imaging Endophenotypes in GWAS, NIH R21AG0570381, 4/1/17-3/31/19, NCE to 3/21.
14. Co-PI, (MPIs: X Shen, W Pan), New Machine Learning Tools for Biomedical Data, NIH RO1 GM081535, 07/11-06/15, NCE 07/15-06/17.

Mentoring Graduate Students and Post-docs

1. Primary mentor of 8 Post-docs.
2. Primary advisor of 20 former PhD students in Biostatistics (all graduated).
3. Project/Plan B advisor of about 20 former MS students in Biostatistics (all graduated).

Teaching: courses taught

1. PubH 7475/8475, Statistical Learning and Data Mining;
2. PubH 7450, Survival Analysis;
3. PubH 7407, Categorical Data Analysis;
4. PubH 5467, Categorical Data Analysis;
5. PubH 5470-2, Statistical Learning and Data Mining;
6. PubH 5470, Statistics in Genetics and Molecular Biology;
7. PubH 8421, Advanced Categorical Data Analysis;
8. PubH 5466, Biostatistical Inference II;
9. PubH 5454, Biostatistics;
10. BINF 5480, Bioinformatics Journal Club.

Service Activity

Professional Service

Fellow, American Statistical Association (ASA).

Fellow, Institute of Mathematical Statistics (IMS).

Member, International Biometric Society (ENAR).

Member, International Genetic Epidemiology Society (IGES).

Member, American Society of Human Genetics (ASHG).

Study Section Member, NIGMS/NIH Biomedical Research and Research Training Review Subcommittee B (BRT-B), Sept 2011–June 2015.

Associate Editor, *Statistics in Biosciences*, Feb 2015–May 2023.

Associate Editor, *Journal of the American Statistical Association*, Oct 2003–2006.

Editorial Board, *Statistics in Medicine*, Aug 2002–November 2005.

Manuscript referee: *Annals of Applied Statistics*, *Applied Statistics*, *Biometrics*, *Biometrical J*, *Biometrika*, *Biostatistics*, *Bernoulli*, *Canadian Journal of Statistics*, *Communications in Statistics*, *Computational Statistics and Data Analysis*, *Controlled Clinical Trials*, *Journal of Biopharmaceutical Statistics*, *Journal of Computational and Graphical Statistics*, *Journal of Multivariate Statistics*, *Journal of Nonparametric Statistics*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society–B*, *Lifetime Data Analysis*,

Scandinavian Journal of Statistics, *Statistica Sinica*, *Statistical Analysis and Data Mining*, *Statistics and Its Interface*, *Statistics in Bioscience*, *Statistics in Medicine*, *Technometrics*, *American Journal of Epidemiology*, *American Journal of Human Genetics*, *American Journal of Public Health*, *Annals of Human Genetics*, *Behavioral Genetics*, *Bioinformatics*, *BMC Bioinformatics*, *Genetic Epidemiology*, *Genome Biology*, *Human Genetics*, *Human Heredity*, *Human Molecular Genetics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *JAMA*, *Nature*, *Nucleic Acids Research*, *Pharmacogenomics*, *PLoS Computational Biology*, *PLoS Genetics*, *PLoS One*, *PNAS*, etc.

Grant proposal review: Army Office of Research, NIH, NSF, Research Grants Council of Hong Kong, University Grants Committee of Hong Kong, Alberta Heritage Foundation for Medical Research (Canada).