# Curriculum Vitae

# Tian Zheng

November 3, 2026

Department of Statistics
Columbia University
Room 1004, MC 4690
1255 Amsterdam Ave
New York, New York 10027
Phone: (212) 851-2149 Fax: (212) 851-2164
Email: tzheng@stat.columbia.edu Homepage: http://www.stat.columbia.edu/∼tzheng

---

## Research Interests

Methodological research for exploring and analyzing interesting patterns in complex data from different application domains.

**Statistics:** applied statistics, Bayesian modeling, metric learning, sparse learning. **Data Science:** modeling of (social) network structure, modeling of dynamic social interactions, design and analysis of sample surveys, dynamic latent space models for evolving relational data, models for high-dimensional spatial-temporal data, applied machine learning for image and LiDAR data, **agentic AI for data science workflows**. **Data science and statistical education: AI education**, **AI literacy for faculty development**. **Statistical genetics:** association mapping of complex traits using high-throughput genomic data. **Bioinformatics and Computational genomics:** Feature selection and classification for gene expression data, data mining of microarray data, hidden Markov models for DNA mutations and phylogenies inference, evaluation of protein regulatory network.

---

## Education

| | |
|---|---|
| 09/1998–05/2002 | Columbia University, New York, NY<br>Ph.D. in Statistics (2002). M.A. in Statistics (2000).<br>Thesis: Multiple-Marker Screening Approach Towards the Study of Complex Traits in Human Genetics. Advisor: Professor Shaw-Hwa Lo. |
| 09/1994–07/1998 | Tsinghua University, Beijing, China<br>B.S. in Applied Mathematics (1998). Minor in Computer Science (1998). Thesis: Mathematical Modeling and Parametrization of a Class of Consonants. |

---

## Academic Appointments

| | |
|---|---|
| 07/2017–present | Professor, Department of Statistics, Columbia University, New York, NY. |
| 10/2021–10/2026 | Deputy Director (2024–2026), Chief Convergence Officer and Education Director (2021–2026), Center for Learning the Earth with Artificial Intelligence and Physics (LEAP), an NSF STC, Columbia University |
| 07/2019–06/2025 | Chair, Department of Statistics, Columbia University, New York, NY |

| | |
|---|---|
| 08/2022-08/2023 | Director of Engineering Workforce Development, Center for Smart Streetscape (CS3), an NSF ERC, Columbia University, New York, NY |
| 07/2017–06/2020 | Associate Director for Education, Data Science Institute (DSI), Columbia University, New York, NY. |
| 12/2025–present | Founder and co-Chair, University Seminar on AI for Social Good and Society, Columbia University, New York, NY |
| 02/2015–present | Affiliate member, Columbia Population Research Center (CPRC), Columbia University, New York, NY |
| 11/2012–present | Associate Member, Herbert Irving Comprehensive Cancer Center (HICCC), Columbia University Medical Center, New York, NY |
| 05/2012–06/2017 | Associate professor (w./ tenure), Department of Statistics, Columbia University, New York, NY. |
| 07/2007–05/2012 | Associate professor (w/o tenure), Department of Statistics, Columbia University, New York, NY. |
| 07/2002–06/2007 | Assistant professor, Department of Statistics, Columbia University, New York, NY. |

---

## University Committees

| | |
|---|---|
| 07/2019–present | Member, Faculty Advisory Committee, Columbia Beijing Global Center |
| 09/2020–present | Member, Provost's Faculty Committee on Educational Innovation, Columbia University, New York, NY |
| 2026–present | Member, University Senate, Columbia University (A&S Tenured Faculty Representative from Natural Sciences) |
| 2026–present | Elected representative of Natural Sciences, Policy and Planning Committee (PPC), Arts and Sciences, Columbia University, New York, NY |
| 09/2014-09/2025 | Member, Shared Research Computing Policy Advisory Committee (SRCPAC); Member (2018 - 2019), training subcommittee of SRCPAC. Columbia University, New York, NY |
| 10/2023–05/2025 | Chair, Steering Committee on Curricular Governance, Arts and Sciences, Columbia University, New York, NY |
| 01/2019–01/2022 | Member, Budget Committee, Arts and Sciences, Columbia University, New York, NY |
| 09/2021–08/2023 | Member (as NS chair representative), Policy and Planning Committee, Arts and Sciences, Columbia University, New York, NY |
| 09/2018–2026 | Co-Chair (2018–2021, 2025–2026), Member (2018–2026), Data Science Education working group, DSI, Columbia University, New York, NY. |
| 12/2020–06/2021 | Member, Search Committee for the Dean of School of Professional Studies, Columbia University |
| 09/2019–05/2021 | Member, Graduate Equity Initiative, Arts and Sciences, Columbia University, New York, NY |
| 05/2013–06/2020 | Chair (2017-2020) and Member, Education Committee, Data Science Institute (DSI), Columbia University, New York, NY. |

| | |
|---|---|
| 09/2017–06/2020 | Member, Research Subcommittee of the Provost's Advisory Committee on the Libraries, Columbia University, New York, NY |
| 09/2019–12/2019 | Member and Subcommittee Chair (computational and theoretical sciences), Science Plan Committee, Arts and Sciences, New York, NY |
| 09/2018–06/2019 | Member, Academic Review Committee (ARC), Arts and Sciences, Columbia University, New York, NY. |
| 09/2018–02/2019 | Member, Uris Vision Committee, Arts and Sciences, Columbia University, New York, NY. |
| 01/2017–10/2017 | Member, Natural Science Equity Committee, Arts and Sciences, Columbia University, New York, NY. |
| 11/2012–07/2017 | Co-Chair (2014-2017) and committee member, Center on Foundations of Data Science, Data Science Institute (DSI), Columbia University, New York, NY. |
| 09/2015–06/2017 | Member, Data and Society Task Force, Columbia University |

---

## Honors

| | |
|---|---|
| 2025-present | Provost Senior Leadership Fellow, Office of Vice Provost for Faculty Advancement, Columbia University |
| 2025 | Elected fellow of Asian American Scholar Forum (AASF) |
| 2024 | Elected fellow of American Association for the Advancement of Science (AAAS) |
| 2022 | Elected fellow of Institute of Mathematical Statistics (IMS) |
| 2022-2023 | Ivy+ Provost Leadership Fellows, Ivy+ Faculty Advancement Network |
| 2021 | Lenfest Distinguished Columbia Faculty Award, Columbia University |
| 2020 | Provost's Senior Faculty Teaching Scholar, Columbia University |
| 2019 | Service Award, American Statistical Association |
| 2019-2021 | Provost Leadership Fellow, Columbia University |
| 2019 | MA Convocation Speaker, Columbia University |
| 2017 | Presidential Teaching Award, Columbia University |
| 2014 | Elected fellow of American Statistical Association |
| 2011 | Elected member of International Statistical Institute. |
| 2010 | Google Faculty Research Award[1], Google, Inc. |
| 2009 | JSM 2009 Poster Award, Section on Survey Research Methods (SRMS) of the American Statistical Association |
| 2008 | The Mitchell Prize[2], Section on Bayesian Statistical Science (SBSS) of the American Statistical Association and the International Society for Bayesian Analysis (ISBA) |

[1]From Google Research University Awards website: "the Google Research Awards program aims to identify and support world-class, full-time faculty pursuing research in areas of mutual interest."

[2]From ISBA website: "The Prize is awarded in recognition of an outstanding paper that describes how a Bayesian analysis has solved an important applied problem. The Prize is jointly sponsored by the Section on Bayesian Statistical Science (SBSS) of the ASA, the International Society for Bayesian Analysis (ISBA), and the Mitchell Prize Founders' Committee."

| | |
|---|---|
| 2008 | Outstanding Statistical Application Award[3], American Statistical Association |
| 2007 | Achievement award, WorldComp'07 (The 2007 World Congress in Computer Science, Computer Engineering and Applied Computing). |
| 2007 | Worldcomp'07 honorary first runner-up for best paper award in Bioinformatics and Computational Biology. |
| 1998 | Graduate of honors with distinction (awarded to top 2% of the graduating class), Tsinghua University. |

---

## Selected Research Grants

Grants for which participating as senior investigator are listed.[4]

| | |
|---|---|
| 2021-2031 | *NSF STC: Center for Learning the Earth with Artificial Intelligence and Physics (LEAP)* PI. P. Gentine.<br>National Science Foundation (NSF) (Role: co-**PI** as Education Director. Total award: $46,700,000.) |
| 2024-2027 | Valuation of Surface Ocean pCO2 Observations For Machine Learning Applications. PI: G. McKinley.<br>National Oceanic and Atmospheric Administration (NOAA) (Role: co-**PI**. Total award: $573,955.) |
| 2022-2028 | *NSF Engineering Research Center for Smart Streetscapes (CS3)* PI. A. Smyth.<br>National Science Foundation (NSF) (Role: co-**PI** as Engineering Workforce Development Director. Total award: $26,000,000.) |
| 2020-2023 | *Combining data and models for robust estimates of global air sea CO2 fluxes* PI. G. McKinley.<br>National Oceanic and Atmospheric Administration (NOAA) (Role: co-**PI**. Total award: $447,280.) |
| 2019-2024 | *Big data methods for comprehensive similarity based risk prediction.* PI: S. Wang.<br>National Institutes of Health, (Role: co-investigator, one-month effort. Total award: $2.236.063.) |
| 2021, 2022 | *Columbia-Lehman Public Interest Technology Data Science Corps (PIT-DSC)* PI: T Zheng.<br>Public Interest Technology University Network Challenge Fund (Role: **PI**. Total award: $180,000 $\times 2$)<br>PIT Data Science Corps is a summer REU program, a collaboration between Columbia University and Lehman College that connects public interest projects in NYC's underserved communities with NYC's own data science talent while diversifying the data science pipeline. |
| 2018-2021 | *BIGDATA: F: Statistical Foundation of Predictivity: A Novel Architecture for Big Data Learning.* PI: S-H Lo.<br>National Science Foundation (Role: co-**PI**, 0.8 month effort. Total award: $900,000.) |

[3]From award website: "The Outstanding Statistical Application Award was established in 1986 to recognize the authors of papers that demonstrate an outstanding application of statistics in any substantive field."

[4]Also participated in a number of training grants and center grants at Columbia University.

2018 *AI for Earth - Keeping a Close Watch on Our Trees: Large-scale Forest Ecological Surveys via a Data Science Workflow using High-Resolution Imaging and Remote Sensing Data.* PI: T. Zheng.

AI for Earth Microsoft Azure Grant and Esri licenses award (Role: **PI**. Total award: $10,000 Azure credit).

2016-2018 *Dissecting Spatiotemporal Brain Dynamics.* PIs: T Zheng and E Hillman.

Data Science Interdisciplinary ROADS Grant, Columbia University

2016-2018 *Reproducible And Scalable Bayesian Modeling For Next-Generation Social Sciences*, PI: T Zheng.

Defense Advanced Research Projects Agency (DARPA), two-month effort. Total projected award: $1,501,742.

2015-2018 *A Novel Statistical Framework for Big Data Prediction*, PI: S-H Lo

National Science Foundation (Role: co-**PI**, one-month effort. Total award: $300,000.)

2010-2015 *Latent Space Models for Aggregated Relational Data in Social Sciences*, PI: T Zheng.

National Science Foundation (Role: **PI**, Total award: $350,000)

2010 *Dynamic Projection State-Space Model for Evolving Relational Data*, PI: T Zheng.

Google Research Award, (Role: **PI**, unrestricted gift $70,000).

2009-2010 *Latent Structure Models for Social Networks Using Aggregated Relational Data*, PI: T. Zheng.

Columbia Population Research Center, (Role: **PI**, $6,000.)

2009-2011 *Computational Discovery of Synergistic Mechanisms Responsible for Psychiatric Disorders*, PI: D Anastassiou and M Karayiorgou.

National Institutes of Health, (Role: co-investigator, one-month effort.)

2009–2011 ARRA Administrative supplement to *"New Approach Towards Mapping of Complex Human Disorders"*, PI: S-H Lo.

National Institutes of Health, (Role: co-**PI**, two-month effort. Total award: $255,157.)

2007–2010 *A General Framework for High Throughput Biological Learning: Theory Development and Applications*, PI: S-H Lo.

National Science Foundation (Role: co-**PI**, one-month effort. Total award: $270,000.)

2006–2010 *Computer System for Functional Analysis of Genomic Data*, PI: A Rzhetsky.

National Institutes of Health R01 grant, (Role: co-investigator, 10% effort.)

2005–2011 *New Approach Towards Mapping of Complex Human Disorders*, PI: S-H Lo.

National Institutes of Health R01 grant, (Role: co-**PI**, 16.7% effort. Total award: $795,703)

2005–2008 *Design and Analysis of "How many X's do you know" surveys for the study of polarization in social networks*, PI: A Gelman.

National Science Foundation (Role: co-**PI**, one-month effort. Total award: $600,000.)

## Peer-reviewed Publications

(Note: Students are in *italic*.)

115. Donoho, D. L., Kang, J., Lin, X., ..., Xing, E. P., Zheng, T., and Zhu, H. (2026). "Rebuilding" Statistics in the Age of AI: Culture, Infrastructure, and Training. *Journal of the American Statistical Association*. In press. arXiv:2601.17510.

114. *Hou, Z.*, Sun, J., *Venkatasubramanian, S.*, *Jin, P.*, Li, S., and Zheng, T. (2025). Calibrating Geophysical Predictions under Constrained Probabilistic Distributions. arXiv:2512.03081. In Revision.

113. Lin, J., Hu, Z., Beucler, T., Frields, K., Christensen, H., Hannah, W., ... and Pritchard, M. (2026). Crowdsourcing the Frontier: Advancing Hybrid Physics-ML Climate Simulation via a $ 50,000 Kaggle Competition. *Journal of Advances in Modeling Earth Systems*, **18**(5), e2025MS005643.

112. Collard, J., Gentine, P., and Zheng, T. (2025). Power Ensemble Aggregation for Improved Extreme Event AI Prediction. arXiv preprint arXiv:2511.11170.

111. Zhong, C., Ji, S., and Zheng, T. (2024). Graph-Enabled Fast MCMC Sampling with an Unknown High-Dimensional Prior Distribution. arXiv preprint arXiv:2408.02122.

110. Ward, O., Smith, A. L., and Zheng, T. (2026) Bayesian Modeling for Aggregated Relational Data: A Unified Perspective. *Social Networks*, **87**, 61–76.

109. *Jaber, A.*, *Zhu, W.*, Roy, A., Agonafir, C., Hawkins, L., Jayavelu, K., Downes, J., Mohamed, S., and Zheng, T. (2026). AutoClimDS: Climate Data Science Agentic AI–A Knowledge Graph is All You Need. *2026 IEEE Conference on Artificial Intelligence (CAI)*, 1057–1063. arXiv:2509.21553.

108. Zheng, T.., *Venkatasubramanian, S.*, Li, S., Braverman, A., *Ke, X.*, *Hou, Z.*, *Jin, P.*, and *Agrawal, S. S.* (2026). Machine Learning Workflows in Climate Modeling: Design Patterns and Insights from Case Studies. arXiv preprint arXiv:2510.03305. *Philosophical Transactions of the Royal Society A*. Accepted.

107. Han, T., Ferraz, A., Fang, D., Zheng, T., Ankori-Karlinsky, R., Arellano, G., ... and Xu, X. (2026). Taller Trees Experienced Less Crown Damage During a Severe Hurricane in a Tropical Forest. *Global Change Biology*, **32**(1), e70709.

106. Boeschoten, L. E., Ankori-Karlinsky, R., Arellano, G., Brown, A. J., Fang, D., Leitold, V., Morton, D., *Yuan, G.*, Zheng, T., Zimmerman, J. K., ... (2026). Tree Growth After a Major Hurricane Reflects Predisturbance Vigor Rather Than Canopy Damage. *Proceedings of the National Academy of Sciences*, **123**(16), e2532451123.

105. El-Bassel, N., David, J., Mukherjee, T. I., Aggarwal, M., Wu, E., Gilbert, L., ... and Zheng, T. (2025). The Practical, Robust Implementation and Sustainability (PRISM)-capabilities model for use of Artificial Intelligence in community-engaged implementation science research. *Implementation Science*, **20**(1), 37.

104. Agonafir, C. and Zheng, T. (2026) Structured Exploration of Machine Learning Model Complexity for Spatio-Temporal Forecasting of Urban Flooding. *Urban Water Journal*, **23**(3), 380–407.

103. Yu, S., Hu, Z., Subramaniam, A., Hannah, W., Peng, L., Lin, J., ... Zheng, T., ..., and Pritchard, M. (2025). ClimSim-Online: A Large Multi-Scale Dataset and Framework for Hybrid Physics-ML Climate Emulation. *Journal of Machine Learning Research*, **26**(142), 1-85.

102. Li, S., Zheng, T., Farchi, A., Bocquet, M.,and Gentine, P. (2025). Probabilistic data assimilation for ensemble distribution projections with generative machine learning: A Lorenz'96 proof-of-concept. *Geophysical Research Letters*, **52**(12), e2024GL112523.

101. Ankori-Karlinsky, R., Jackson, T. D., *Yuan, G.*, Zimmerman, J. K., Morton, D. C., Zheng, T., and Uriarte, M. (2025). Wind acclimation in a subtropical forest: trees on wind-exposed slopes are shorter with smaller crowns. *New Phytologist*.

100. Lin, X., Cai, T., Donoho, D., Fu, H., Ke, T., Jin, J., ... and Zhu, J. (2025). Statistics and AI: A fireside conversation. *Harvard Data Science Review*, **7**(2).

99. Zheng, T., Keyes, K., Ji, S., Calderon, A., Wu, E., Doogan, N. J., ... and Feaster, D. J. (2025). Opioid use disorder prevalence in 57 New York counties from 2017 to 2019: A Bayesian evidence synthesis. *Drug and Alcohol Dependence*, **267**, 112548.

98. *Jones, T. D.*, *Ward, O. G.*, *Jiang, Y.* and Zheng, T. (2025) Scalable Community Detection in Massive Networks using Aggregated Relational Data. *Statistical Sinica* **35**, 1-23

97. *Lu, D.*, Smith-Martin, C., Muscarella, R., Uriarte, M., Zheng, T. (2025) A Spatio-temporal Model of Embolism Propagation in Leaf Vein Networks. *AoB Plants.*

96. El-Bassel, N., David, J. L., Aragundi, E., Walters, S. T., Wu, E., Gilbert, L., ... and Zheng, T. (2025). Artificial Intelligence and Stigma in Addiction Research: Insights From the HEALing Communities Study Coalition Meetings. *Journal of Addiction Medicine*, **20**(2), 198.

95. Li, S., Qu, Y., Zheng, T. , and Gentine, P. (2024) A Machine-Assisted Physical Closure for Suspended Sediments in Vegetated Turbulent Channel Flows. *Geophysical Research Letters*, **51** (20), e2024GL110475

94. *Kim, S.*, *Nathaniel, J.*, *Hou, Z.*, Zheng, T., and Gentine, P. (2024) Spatiotemporal Upscaling of Sparse Air-Sea pCO21 Data via Physics-Informed Transfer Learning. *Scientific Data*, **11**(1), 1098.

93. Yu, S., Hannah, W. M., Peng, L., Bhouri, M. A., Gupta, R., Lin, J., ... and Pritchard, M. S. (2023). ClimSim: An open large-scale dataset for training high-resolution physics emulators in hybrid multi-scale climate simulators. arXiv:2306.08754. *Advances in Neural Information Processing Systems*, **36**.

92. *Fang, G.*, *Ward, O. G.*, and Zheng, T. (2024) Online Community Detection for Event Streams on Networks. arXiv preprint arXiv:2009.01742. *Statistics and Computing* **34**:35

91. Papachristodoulou, A., Heidegger, I., Virk, R. V., et al. (2023) Metformin Overcomes the Consequences of NKX3. 1 Loss to Suppress Prostate Cancer Progression. *European urology.*

90. Uriarte, M., *Tang C.*, Morton, D. C., Zimmerman, J. K., Zheng, T. (2023) 20th-Century hurricanes leave long-lasting legacies on tropical forest height and the abundance of a dominant wind-resistant palm. *Ecology and Evolution* **13**:11.

89. *Tang, C.*, *Lessen, N.*, Wei, Y. and Zheng, T. (2023) Wasserstein Distributional Learning via Majorization-Minimization. *The 26th International Conference on Artificial Intelligence and Statistics.* PMLR 206:10703-10731.

88. Broderick, T., Gelman, A., Meager, R., Smith, A. L., and Zheng, T. (2023). Toward a Taxonomy of Trust for Probabilistic Machine Learning. arXiv preprint arXiv:2112.03270. *Science Advances.* **9**:7.

87. Smith, A. L., Zheng, T., and Gelman, A. (2023). Prediction scoring of data-driven discoveries for reproducible research. *Statistics and Computing*, **33**(1), 11.

86. *Ward, O. G.*, *Wu, J.*, Zheng, T. and Smith, A. L.,. (2022) Network Hawkes Process Models for Exploring Latent Hierarchy in Social Animal Interactions. arXiv preprint arXiv:2012.09598. *Journal of the Royal Statistical Society: Series C (Applied Statistics).* **71**(5): 1402-1426.

85. Liu, L., Meng, Y., Wu, X., Ying, Z. and Zheng, T.[6] (2022). Log-Rank-Type Tests for Equality of Distributions in High-Dimensional Spaces. *Journal of Computational and Graphical Statistics*, **31**(4), 1384-1396.

84. *Wu, J.*, *Ward, O. G.*, Curley, J., and Zheng, T.[6] (2022) Markov-Modulated Hawkes Processes for Sporadic and Bursty Event Occurrences. *Annals of Applied Statistics,* **16** (2), 1171-1190.

83. Gloege, L., Yan, M., Zheng, T., and McKinley, G. (2022) Improved quantification of ocean carbon uptake by using machine learning to merge global models and pCO2 data. *Journal of Advances in Modeling Earth Systems*, **14**(2), e2021MS002620.

82. *Palmer, W.,* Davis, R., and Zheng, T. (2021) Count-Valued Time Series Models for COVID-19 Daily Death Dynamics. *Stat,* **10**(1), e369.

81. *Tang C.*, Uriarte, M., *Jin, H.*, Morton, D. and Zheng, T. (2021) Large-Scale, Image-Based Tree Species Mapping in a Tropical Forest using Artificial Perceptual Learning. *Methods in Ecology and Evolution,* **12**(4), 608-618

80. Zaugg, I., Culligan, P. J., Witten, R., and Zheng, T.[6] (2021) Collaboratory at Columbia: An Aspen Grove of Data Science Education. *Harvard Data Science Review,* **3**(4).

79. Guo, J., Yuan, C., Ning, S., Zheng, T., Bello, N., Kiryluk, K., and Weng, C. (2021) Similarity-based Health Risk Prediction Using Domain Fusion and Electronic Health Records Data. Journal of Biomedical Informatics. **116**, 103711.

78. *Ward, O.G.*, *Huang, Z.*, *Davison, A.* and Zheng, T.[6] (2021). Next waves in veridical network embedding. Statistical Analysis and Data Mining: The ASA Data Science Journal. **14**(1): 5-17

77. *Palmer, W. R.* and Zheng, T. (2020) Spectral Clustering in Directed Networks. International Workshop on Complex Networks and their Applications 2020

76. *Wu, J.*, Smith, A. L., and Zheng, T.[6] (2019) Diagnostics and Visualization of Point Process Models for Event Times on a Social Network. Proceedings of the 18th Conference of the Applied Stochastic Models and Data Analysis International Society.

75. *Zhu, K.*, *Yang, T. H. O.*, Dorie, V., Zheng, T., and Anastassiou, D. (2019). Meta-analysis of expression and methylation signatures indicates a stress-related epigenetic mechanism in multiple neuropsychiatric disorders. Translational psychiatry, **9**(1), 32.

74. *Sahai, S.*, *Jones, T.*, Cowan, S. K., and Zheng, T.[6] (2018, December). Estimating Personal Network Size with Non-random Mixing via Latent Kernels. In International Workshop on Complex Networks and their Applications (pp. 694-705). Springer, Cham.

73. *Hsu, Y.*, *Auerbach, J.*, Zheng, T. and Lo, S.H. (2018+). Coping with Family Structure in Genome Wide Association Studies: a Comparative Evaluation. In BMC proceedings (Vol. 12, No. 9, p. 42). BioMed Central.

72. Tan, S., Makela, S., Heller, D., Konty, K., Balter, S., Zheng, T. , and Stark, J. H. (2018+). A Bayesian Evidence Synthesis Approach to Estimate Disease Prevalence in Hard-To-Reach Populations: Hepatitis C in New York City. arXiv preprint arXiv:1801.04587. *Epidemics*, **23**, 96-109.

71. *Meng, L.* and Zheng, T. (2017+). Phase-Aligned Spectral Filtering for Decomposing Spatiotemporal Dynamics. *arXiv preprint*: arXiv:1604.04899. Under revision.

70. *Ma, Y.*, *Ding, Y.* and Zheng, T. (2018). Feature subspace learning based on local point processes patterns. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **11**(1), pp.32-50.

69. *Ma, Y.* and Zheng, T. (2018). Stabilized Sparse Online Learning for Sparse Data. *arXiv preprint*: arXiv:1604.06498. *Journal of Machine Learning Research*, **18** (1), 4773-4808.

68. *Liu Y.*, de la Pena, V. and Zheng T.[6] (2018) Kernel-based measures of association. *WIREs Computational Statistics* **10** (2), e1422.

67. Owczarek *et. al.* (2017) ARF Confers a Context-Dependent Response to Chemotherapy in Muscle-Invasive Bladder Cancer. *Cancer research*, **77**(4), 1035-1046.

66. Lo, A., Chernoff, H., Zheng, T., and Lo, S.-H. (2016). Framework for making better predictions by directly estimating variables' predictivity. *Proceedings of the National Academy of Sciences* **113**(50) 14277-14282.

65. Kalinsky et al (2016) Proteomic Modulation in Breast Tumors after Metformin Exposure: Results from a "Window of Opportunity" Trial. *Clinical and Translational Oncology*, 1-9.

64. He, R. and Zheng, T. (2016). Estimating exponential random graph models using sampled network data via graphon. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

63. *Lo, A.*, *Agne, M.*, *Auerbach, J.*, Fan, R., Lo, S.-H., Wang, P. and Zheng, T.[6] (2016) Network-guided interaction mining for the blood pressure phenotype of unrelated individuals in GAW 19. *BMC Proceedings for GAW 19*. BMC proceedings **10** (7), 333.

62. *Auerbach, J.*, *Agne, M.*, Fan, R., *Lo, A.*, Lo, S.-H., Zheng, T. and Wang, P. (2016) Identification regions of disease related variants in admixed populations with summation partition approach. *BMC Proceedings for GAW 19*. BMC proceedings **10** (7), 131.

61. Zhang et al. (2016) Nkx3.1 Controls the DNA Repair Response in the Mouse Prostate. Prostate **76** (4): 402-8.

60. *Ma Y.*, and Zheng T.[6] (2015+) sDist: Adaptive Sparse Distance Metric Learning. *Statistical Data Analysis and Mining*. In press.

59. Tan, L.S.L, Chan, A. H., and Zheng T.[6] (2015) Topic-adjusted visibility metric for scientific articles. *Annals of Applied Statistics* **10**(1): 1-31.

58. Davis, R.A., *Zang, P.*, and Zheng, T. (2016) Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*. **25**(4): 1077-1096.

57. *Lo, A.*, Chernoff, H., Zheng, T. and Lo, S.-H. (2015) Why aren't significant variables automatically good predictors. *Proc Natl Acad Sci U S A* **112** (45):13892-13897.

56. *Liu, Y.*, Gelman, A. and Zheng, T. (2015) Simulation-Efficient Shortest Probability Intervals. *Statistics and Computing*.**25**(4): 809-819.

55. Cai, B., Zheng, T. and Gelmann E. P. (2015) NKX3.1 Suppresses TMPRSS2-ERG Gene Rearrangement and Mediates Repair of Androgen Receptor-Induced DNA Damage. *Cancer Research* **75**(13): 2686-2698.

54. *He R.* and Zheng T.[6] (2015) GLMLE: Graph-limit Enabled Fast Computation for Fitting Exponential Random Graph Models to Large Social Networks. *Social network analysis and mining*, **5** (1): 1:19.

53. *McCormick, T. H.* and Zheng, T. (2015) Latent space models for aggregated relational data. *Journal of American Statistical Association*. **110** (512): 1684-1695.

52. Avagyan S. *et al.* (2014) Hematopoietic stem cell dysfunction underlies the progressive lymphocytopenia in XLF/Cernunnos deficiency. *Blood*, **124** (10): 1622-1625.

51. Ou Yang, T.-H., Cheng, W.-Y., Zheng, T., Maurer, M. A., and Anastassiou, D. (2014) Breast cancer prognostic biomarker using attractor metagenes and the FGD3-SUSD3 Metagene. *Cancer Epidemiology Biomarkers & Prevention* 0399.2014.

50. Aytes A., Mitrofanova A., Lefebvre C., Alvarez M. J., Castillo-Martin M., Zheng T., Eastham J. A., Gopalan A., Pienta K. J., Shen M. M., Califano A. and Abate-Shen C. (2014) Global analyses of cross-species regulatory networks identifies FOXM1 and CENPF as synergistic master regulators of prostate cancer malignancy. *Cancer Cell* **25**(5): 638-651.

49. McCormick, T. H. and Zheng, T. (2014) Network-based methods for accessing hard-to-reach populations using standard surveys. In *Handbook on Hard-to-Survey Populations.* (ed. Tourangeau, R., Edwards, E., Johnson, T. P., Wolter, K. M., and Bates, N.) Cambridge University Press.

48. *He R.* and Zheng T.[6] (2013) Estimation of exponential random graph models for large social networks via graph limit. *Proceedings of The 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013).* pp248-255. Accepted as "Full Paper." (13% acceptance rate).

47. *Agne M.*, Huang C.-H., Hu I., Wang H., Zheng T. and Lo S.-H. (2013) Considering interactive effects in the identification of influential regions in extremely rare variants via fixed bin approach. *BMC Proceedings*, **8**(S1): S7.

46. Wang H., Huang C.-H., Zheng T., Lo S.-H. and Hu I. (2013) Discovering pure gene-environment association in blood pressure GWAS data: a two-step approach incorporating a new statistic. *BMC Proceedings*, **8**(S1): S62.

45. *Fan R.*, *Huang C. H.*, Hu I., Wang H., Zheng T., Lo S.-H. (2013) A partition-based approach to identify gene-environmental interactions in Genome-Wide Association Studies. *BMC Proceedings* **8**(S1): S60.

44. *Liu Y.*, *Huang C. H.*, Hu I., Lo S.-H., and Zheng, T[6] (2013) A dual clustering framework for association screening with whole genome sequencing data and functional traits. *BMC Proceedings* **8**(S1): S47 (correction S112).

43. Irshad, S., Bansal, M., Castillo-Martin, M., Zheng, T., Aytes, A., Wenske, S., Magnen, C.L., Guarnieri, P., Sumazin, P., Benson M.C., Shen, M.M., Califano, A., and Abate-Shen, C. (2013) A molecular signature predictive of indolent prostate cancer. *Science Translational Medicine* **5** (202): 202ra122.

42. *Zhou, H.*, Zheng, T.[6] (2013) Bayesian Hierarchical Graph-Structured Model for Pathway Analysis Using Gene Expression Data. *Statistical Applications in Genetics and Molecular Biology* **12**(3):. 393-412

41. McCormick, T. H., *Moussa, A.*, *Ruf, J.*, DiPrete, T. A., Gelman, A., Teitler, J., Zheng, T.[6] (2013) Measuring social distance using indirectly observed network data. *Journal of Statistical Theory and Practice* **7**(1): 120-132.

40. Wang, H. Lo, S.-H., Zheng, T. and Hu, I. (2012) Interaction-based feature selection and classification for high-dimensional biological data. *Bioinformatics* **28**(21)2834-2842.

39. McCormick, T. H., and Zheng, T. (2012) Latent demographic profile estimation in at-risk populations. *Annals of Applied Statistics* **6**(4):1795-1813.

38. *McCormick, T. H.*, *He, R.*, Kolacyzk, E. D. and Zheng, T.[6] (2012) Surveying hard-to-reach groups through sampled respondents in a social network. *Statistics in Biosciences* **4**(1):177-195

37. *McCormick, T. H.*, Zheng, T. and Salganick, M. J. (2011). B: ORGANIZATIONAL BEHAVIOR-How Many People Do You Know? Efficiently Estimating Personal Network Size Bb: 140. *Operations Research Management Science***51**(1): 49.

36. *Wang, H.*, *Huang, C.-H.*, Lo, S.-H., Zheng, T., and Hu, I. (2011) New insights on old methods in identifying causal rare variants. *BMC Proceedings* for the Genetic Analysis Workshop 17 (Boston, MA) **5**(s9):s50.

35. *Agne, M.*, *Huang, C.-H.*, Hu, I., *Wang, H.*, Zheng, T., and Lo, S.-H. (2011) Identifying Influential Regions in Extremely Rare Variants using a Fixed-Bin Approach. *BMC Proceedings* for the Genetic Analysis Workshop 17 (Boston, MA) **5**(s9):s3.

34. *Liu, Y.*, *Huang, C.-H.*, Hu, I., Lo, S.-H., and Zheng, T.[6] (2011) Association Screening for Genes with Multiple Potentially Rare Variants: an Inverse-Probability Weighted Clustering Approach. *BMC Proceedings* for the Genetic Analysis Workshop 17 (Boston, MA) **5**(s9):s106.

33. *Fan, R.*, *Huang, C.-H.*, Lo, S.-H., Zheng, T., and Ionita-Laza, I. (2011) Identifying Rare Disease Variants in the Genetic Analysis Workshop 17 Simulated Data: A Comparison of Several Statistical Approaches. *BMC Proceedings* for the Genetic Analysis Workshop 17 (Boston, MA) **5**(s9):s17.

32. Zheng, T., Chernoff, H., Hu, I., Ionita-Laza, I., Lo, S.-H. (2011) Discovering influential variables: a general computer intensive method for common genetic disorders. In Lu, H.H.S., Schölkopf, B., Zhao, H. (Eds.) (2011) Handbook of Statistical Bioinformatics, Springer-Verlag, New York. MS Link.

31. DiPrete, T. A., *McCormick, T. H.*, Gelman, A., Teitler, J., Zheng, T. (2011) Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology* **116** (4):1234-83

30. Kim, S., Singh, P., Park, J., Park, S., Friedman, A., Zheng, T., Sun, Q., Lee, Y.-H. and Lee, K. (2011) Fungal photoreceptor MGWC-1–mediated disease suppression in rice. *Fungal Genetics and Biology* **48**(4):400-407.

29. Zheng, T. and Gastwirth, J. L. (2011) On Bootstrap Tests of Symmetry About an Unknown Median. *Journal of Data Science* **8**(3): 397-412

28. Chernoff, H., Lo, S. H. and Zheng, T. (2009) Discovering influential variables: a method of partitions. *Annals of Applied Statistics* **3** (4):1335-1369.

27. *McCormick, T. H.*, Salganik, M. J., and Zheng, T. (2010) How many people do you know?: efficiently estimating personal network size. *Journal of the American Statistical Association* **105**(489):59-70 (Highlighted in the April 2010 issue of Amstat News, the official membership magazine of the American Statistical Association.)

26. Wang, S., Zheng, T., Chanock, S., Jedrychowski, W. and Perera, F. P. (2010) Methods for detecting interactions between genetic polymorphisms and prenatal environment exposure with a mother-child design. *Genetic Epidemiology* **34**(2) 172-180.

25. An, P., Mukherjee, O., Chanda, P., Yao, L., Engelman, C. D., Huang, C.-H., Zheng, T., Kovac, I. P., Dube, M.-P., Liang, X., Li, J., de Andrade, M., Culverhouse, R., Malzahn, D., Manning, A. K., Clarke, G. M., Jung, J., Province, M. A. (2009) The challenge of detecting epistasis (GxG interactions): Genetic Analysis Workshop 16. *Genetic Epidemiology* **33**(s1) s58-s67.

24. *Qiao, B.*, *Huang, C.-H.*, *Cong, L.*, Xie, J., Lo, S.-H., Zheng, T.[5] (2009) Genome-Wide Gene-Based Analysis of Rheumatoid Arthritis-Associated Interaction with PTPN22 and HLA-DRB1. *BMC Proceedings for the Genetic Analysis Workshop 16* **3**(s7):s132.

23. *Huang, C.-H.*, *Cong, L.*, Xie, J., *Qiao, B.*, Lo, S.-H., Zheng, T.[5] (2009) Rheumatoid Arthritis-Associated Gene-Gene Interaction Network for Rheumatoid Arthritis Candidate Genes. *BMC Proceedings for the Genetic Analysis Workshop 16* **3**(s7):s75.

22. Salicru, M., Vives, S. and Zheng, T.[5] (2009) Inferential clustering approach for two-color spotted microarray experiments. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 6(4):594-604. (Featured article from the October-December 2009 issue).

21. *Watkinson, J.*, Liang, K.C., Wang, X., Zheng, T., and Anastassiou, D. (2008) Inference of regulatory gene interactions from expression data using three-way mutual information. *The Annals of New York Academy of Sciences* **1158**:302-313.

20. Lo, S.-H., Chernoff, H., *Cong, L.*, *Ding, Y.* and Zheng, T. (2008) Discovering interactions among BRCA1 and other candidate genes involved in Sporadic Breast Cancer. *Proc Natl Acad Sci U S A* **105**: 12387-12392.

19. *Iossifov, I.*, Zheng, T., Baron, M., Gilliam, T.C. and Rzhetsky, A. (2008) Genetic-linkage mapping of complex hereditary disorders to a whole-genome molecular-interaction network. *Genome Research* **18**:1150-1162.

[5]Joint correspondence author.

18. *Watkinson, J.*, Wang, X., Zheng, T. and Anastassiou, D. (2008) Identification of gene-gene interactions associated with disease from gene expression data. *BMC Systems Biology* **2**:10.

17. *Yan, X.* and Zheng, T.[6] (2008) Selecting informative genes for discriminant analysis using multigene expression profiles. *BMC Genomics* **9**(S2):S14.

16. *Chen, G. K.*, Zheng, T., Witte, J. S. and Goode, E. L. (2007) Genome-wide association analyses of expression phenotypes. *Genetic Epidemiology* **31**(S1):S7-S11.

15. *Woo, J.H.*, Zheng, T.[7] and Kim, J.H. (2007) Identifying genomic regulators of set-wise co-expression. *Proceedings of IEEE 7th International Conference on Bioinformatics and BioEngineering.* paper URL. (Acceptance for regular research papers is 65 out of $\sim$ 500.) Extended paper published in *International Journal of Functional Informatics and Personalised Medicine* **1**:407-418.

14. *Yan, X.* and Zheng, T.[6] (2007) Discriminant Analysis using multigene profiles in classification of breast cancer. *Proceedings of the 2007 International Conference on Bionformatics and Computational Biology.* (Acceptance rate: 30%. Worldcomp'07 Honorary First Runner-Up for Best Paper Award in Bioinformatics and Computational Biology.) Ms file.

13. Zheng, T., Wang, S., *Cong, L.*, *Ding, Y.*, Ionita-Laza, I. and Lo, S.H. (2007) Joint study of genetic regulators for expression traits related to breast cancer. In "Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci." *BMC Proceedings* **1** S1:S10.

12. Wang, S., Zheng, T. and Wang, Y. (2007) Transcription activity hotspot, is it real or an artifact? In "Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci." *BMC Proceedings* **1** S1:S94.

11. Li, Z., Zheng, T., Califano, A. and Floratos, A. (2007) Pattern-based mining strategy for multi-Locus association analysis on GAW15 problem 2 chromosome 18 dataset. In "Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci." *BMC Proceedings* **1** S1:S16.

10. *Ding, Y.*, *Cong, L.*, Ionita-Laza, I., Lo, S.H. and Zheng, T.[6] (2007) Constructing gene association network for rheumatoid arthritis using the backward genotype-trait association (BGTA) Algorithm. In "Genetic Analysis Workshop 15: Gene Expression Analysis and Approaches to Detecting Multiple Functional Loci." *BMC Proceedings* **1** S1:S13.

9. Rzhetsky, A., Wajngurt, D., Park, N. and Zheng, T. (2007) Probing genetic overlap among complex human phenotypes, *Proc Natl Acad Sci U S A* **104**, 11694-11699.

8. Zheng, T., *Ichiba, T.* and Morton, B.R. (2007) Assessing substitution variation across sites in grass chloroplast DNA, *Journal of Molecular Evolution*, **64**, 605-613.

7. *Kerman, J.*, Gelman, A., Zheng, T. and *Ding, Y.* (2007) Visualization in Bayesian data analysis. In Chen, C.-H., Hardle, W. and Unwin, A. (eds), *Handbook of Computational Statistics (Volume III) Data Visualization.* Springer-Verlag, Heidelberg. Paper URL.

6. Zheng, T. and Lo, S.H. (2007) A modified Kendall rank-order association test for evaluating the repeatability of two studies with a large number of objects. In Nair, V. (ed), Advances in Statistical Modeling and Inference—Essays in Honor of Kjell A. Doksum. World Scientific, Hackensack, NJ. Ms file.

5. Rzhetsky, A., Zheng, T.[7] and *Weinreb, C.* (2006) Self-correcting maps of molecular pathways, *PLoS One*, **1**, e61.

[6]Correspondence author.
[7]Joint first author.

4. Zheng, T., *Wang, H.* and Lo, S.H. (2006) Backward genotype-trait association (BGTA)-based dissection of complex traits in case-control designs, *Hum Hered*, **62**, 196-212.

3. Zheng, T., Salganik, M.J. and Gelman, A. (2006) How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks, *Journal of the American Statistical Association*, **101**, 409-423.

2. Lo, S.H. and Zheng, T. (2004) A demonstration and findings of a statistical approach through reanalysis of inflammatory bowel disease data, *Proc Natl Acad Sci U S A*, **101**, 10386-10391.

1. Lo, S.H. and Zheng, T. (2002) Backward haplotype transmission association (BHTA) algorithm—a fast multiple-marker screening method, *Hum Hered*, **53**, 197-215.

---

## Other publications

(Note: Students are in *italic*.)

127. Reeve, J., Zaugg, I. A., Zheng, T., *Yuan, S.*, and *Zhang, Z.* (2021, September 15). Mapping the Field of Data Ethics: A Roadmap for Educators, in *AOIR Ethics Panel 1: Practices and Roadmaps. AoIR Selected Papers of Internet Research. Association of Internet Researchers Conference 2021*, Temple University, Philadelphia, U.S.A. Journal of Information, Communication and Ethics in Society `https://doi.org/10.1108/JICES-12-2021-0124`.

126. *Sun, S.*, *Ward, O. G.*, *Wu, J.*, *Xiao, L.*, *Zhao, X.*, and Zheng, T. (2021). ppdiag: Diagnostic Tools for Temporal Point Processes. *Journal of Open Source Software*, **6**(61), 3133.

125. *Tang, C.*, *Yuan, G.*, and Zheng, T. (2021) Weakly Supervised Learning Creates a Fusion of Modeling Cultures. Invited Discussion for *the 20th anniversary of Breiman et al (2001) "Statistical Modeling: The Two Cultures"*, Observational Studies. *Observational Studies*, **7**(1), 203-211.

124. Zheng, T. and Ying, Z. (2013) Columbia University Statistics. In Agresti, A. and Meng, X.-L. (eds) "*Strength in Numbers: The Rising of Academic Statistics Departments in the U. S.*", Springer, New York.

123. *Huang, C.-H.*, *Agne, M.*, Hu, I., Ionita-Laza, I., Zheng, T., and Lo, S.-H. (2011) GAW 17-Identifying Influential Variables on Unrelated Individual Data. Genetic Analysis Workshop 17 (Boston, MA).

122. *McCormick, T. H.*, and Zheng, T. (2010) A latent space representation of overdisperson in "How many X's do you know?" data. *Proceedings of the joint statistical meeting 2010.*

121. *McCormick, T. H.*, *Moussa, A.*, *Ruf, J.*, *DiPrete, T. A.*, Gelman, A., *Teitler, J.*, Zheng, T. (2009) Comparing two methods for predicting opinions using social structure. *Proceedings of the joint statistical meeting 2009.* (Manuscript #33 above is an extended journal paper version of this report.)

120. *McCormick, T. H.*, Zheng, T. (2009) Towards a unified framework for inference with aggregated relational data. *Proceedings of the joint statistical meeting 2009.* Link

119. Zheng, T. and Lo, S.-H. (2008) Discussion on "Quantifying the Fraction of Missing Information for Hypothesis Testing in Statistical and Genetic Studies by Nicolae, Meng and Kong." *Statistical Science.* **23**(3):321-324.

118. *McCormick, T. H.* and Zheng, T. (2007) Adjusting for recall bias in "how many X's do you know?" surveys, *Proceedings of the joint statistical meeting 2007.* Preprint URL.

117. *Rosner, E.* and Zheng T (2006) Helpfulness of the genders—a student project. STATS. (This is a magazine style article on Teaching Statistics)

116. *Ding, Y.* and Zheng, T. (2006) Tree-Based Integration of One-versus-Some (OVS) Classifiers for Multiclass Classification, *Proceedings of the joint statistical meeting 2006*. Preprint URL.

115. Zheng, T. (2002) Multiple-marker screening approach towards the mapping of complex traits in human genetics. Columbia University Doctoral Dissertation.

---

## Unpublished white paper, technical reports and working papers

(Note: Students are in *italic*.)

138. Balaji, A., Yeltekin, B. D., and Zheng, T. (2026). ForkSCOPE: Charting the Agentic Garden of Forking Paths. arXiv preprint arXiv:2609.12438.

137. Zheng, T. (2026). Probing the Stochastic Machine: Engaging with LLMs in Statistics Curricula Through Veridical Data Science. arXiv preprint arXiv:2606.29754.

136. Zheng, T. and *Hsu, K.-T.* (2026). Grading the Grader: Lessons from Evaluating an Agentic Data Analysis System. arXiv preprint arXiv:2606.24839.

135. *Hou, Z.* and Zheng, T. (2026). Target-Aware Linear Regression Under Distribution Shift. *Journal of Data Science*. Accepted. arXiv:2606.22775.

134. Zheng, T. (2025+). AI Education in Higher Education: A Taxonomy for Curriculum Reform and the Mission of Knowledge. arXiv preprint arXiv:2509.23363.

133. Chernoff, H., Lo, S. H., Zheng, T., and Lo, A. (2017). Estimating the theoretical error rate for prediction. *arXiv preprint* arXiv:1709.02899.

132. Davis, R.A., *Zang, P.*, and Zheng, T. (2015) Reduced-Rank Covariance Estimation in Vector Autoregressive Modeling. *arXiv preprint*: arXiv:1412.2183.

131. *Ding, Y.*, Lo, S.-H. and Zheng, T. (2009) Sum of Squared Local Sums (SSLS): a nonparametric association measure.

130. *Huang, C.-H.*, *Qiao, B.*, *Cong, L.*, Zheng, T. and Lo, S.-H. (2009) Genomewide analysis of genetic interactions with BRCA1 associated with the risk of sporadic breast cancer.

129. Ionita-Laza, I., *Ding, Y.*, *Cong, L.*, Lo, S.H. and Zheng, T. (2007) Combined Linkage and Association Analysis of the NARAC Dataset. Technical report. Department of Statistics, Columbia University.

128. Zheng, T., *Wang, H.* and Lo, S.H. (2004) Rank-based generalized genotype-trait association algorithm for the mapping of quantitative trait loci. Technical report. Department of Statistics, Columbia University.

---

## Selected Invited Seminar Talks

| | |
|---|---|
| July 2026 | Department of Statistics and Data Science, Tsinghua University |
| April 2026 | Department of Statistics, Rice University |
| March 2026 | Department of Biostatistics, University of North Carolina at Chapel Hill |
| Sep 2025 | Department of Statistics and Data Science, Washington University at St. Louis |
| Oct 2024 | Department of Statitics, University of Wisconsin at Madison |

| | |
|---|---|
| May 2024 | Columbia LDEO Seminar |
| October 2023 | Department of Statistics, George Washington University |
| May 2023 | Department of Statistics, Stanford University |
| September 2022 | Department of Statistics, Harvard University |
| March 2021 | Department of Statistics, University of Pennsylvania |
| December 2020 | Pacific Institute for the Mathematical Sciences and Simon Fraser University |
| March 2020 | Probability and Society Symposium, Columbia University |
| October 2019 | Department of Statistics, Duke University |
| June 2019 | Center of Statistical Science, Tsinghua University |
| May 2019 | Department of Statistics, University of California at Irvine |
| October 2018 | Science Research Fellow seminar, Columbia University |
| September 2018 | Department of Biostatistics, Vanderbilt University |
| May 2018 | Mount Sinai School of Medicine |
| October 2017 | Undergraduate Statistics Seminar, Columbia University |
| October 2017 | Science Research Fellow seminar, Columbia University |
| September 2016 | Undergraduate Statistics Seminar, Columbia University |
| September 2016 | Science Research Fellow seminar, Columbia University |
| July 2016 | Tsinghua University Statistics Summer Camp |
| February 2016 | Center for the Promotion of Research Involving Innovative Statistical Methodology (PRISM), New York University |
| July 2015 | Institute of System Sciences, Chinese Academy of Sciences |
| Oct 2014 | Department of Biostatistics, Yale University |
| April 2014 | Microsoft Research NYC |
| April 2014 | Department of Psychiatry, Columbia University |
| Oct 2012 | Johnson Research Labs |
| Oct 2011 | Department of Statistics, University of Illinois at Urbana-Champagne |
| Oct 2011 | Department of Statistics, University of Illinois at Chicago |
| Nov 2010 | Department of Statistics, University of Michigan |
| Oct 2010 | Department of Mathematics and Statistics, Boston University |
| April 2009 | Department of Statistics, Rutgers University |
| April 2008 | Department of Statistics, George Washington University |
| Feb 2008 | Division of Biostatistics, New York University |
| Dec 2007 | Department of Statistics, University of Minnesota |
| Nov 2007 | Department of Statistics, Purdue University |

| | |
|---|---|
| Oct 2007 | Department of Biostatistics, University of Pennsylvania |
| June 2007 | Bioinformatics Division, Department of Automation, Institute of Information Processing, Tsinghua University (Beijing, China) |
| November 2006 | Department of Statistics, Yale University. |
| November 2006 | Department of Statistics, Harvard University. |
| November 2005 | University seminar on Genetic Epidemiology, Columbia University. |
| April 2005 | Department of Biostatistics, Yale University. |
| November 2004 | Department of Statistics, Columbia University. |
| November 2003 | Department of Biostatistics, University of Alabama at Birmingham. |
| October 2002 | International Center for Health Outcomes and Innovation Research, Columbia University. |
| February 2002 | Department of Statistics, University of Wisconsin at Madison. |
| February 2002 | Department of Statistics, University of California at Riverside. |
| February 2002 | Department of Statistics, University of California at Davis. |
| February 2002 | Department of Statistics, University of North Carolina at Chapel Hill. |
| February 2002 | Department of Statistics, Pennsylvania State University. |
| February 2002 | Department of Statistics, Columbia University. |
| February 2002 | Department of Statistics, Ohio State University. |

---

## Selected Invited conference presentations

| | |
|---|---|
| Oct 2026 | Grow 2026, Columbia University – Banquet Speaker |
| Aug 2026 | JSM 2026 – Panelist on NASEM report |
| July 2026 | Keynote speaker, International Forum on Data Science, Renmin University, Beijing, China |
| June 2026 | Panelist on NASEM Report, IMSI, Chicago |
| June 2026 | ICSA Leadership Summit, ICSA 2026, George Mason University, Virginia |
| May 2026 | 1st Workshop: Statistics in Trustworthy AI, IVADO, Montreal, Canada |
| April 2026 | Keynote Speaker and Panelist, University Anniversary Celebration, Department of Statistics and Data Science, Tsinghua University |
| July 2025 | ICSA China, Zhuhai, China – Banquet Speaker |
| March 2025 | ENAR, New Orleans, LA – "PANEL: AI and Statistics: Challenges and Opportunities." |
| December 2024 | ICSDS 2024, Nice, France. "Distribution-Informed Learning Via Kernelized Stein Discrepancy Calibration." |
| November 2024 | Workshop on Extreme Events, University of Chicago. "Distribution-Informed Learning Via Kernelized Stein Discrepancy Calibration." |

| | |
|---|---|
| August 2024 | JSM 2024 – "A Bayesian Information Synthesis Framework for Opioid Use Disorder Prevalence Estimation" |
| August 2024 | 2024 IMS New Researchers Conference – "Statistical Challenges in Climate Data Science" |
| June 2024 | WNAR 2024, Fort Collins, CO |
| June 2024 | Veridical Data Science Workshop, University of California at Berkeley. |
| May 2024 | Statistics in the Age of AI, University of George Washington |
| April 2024 | Keynote Speaker, Prob/Stat Day, UMBC |
| February 2024 | SIAM Uncertainty Quantification (Trieste, Italy) 2024 (Plenary) |
| June 2023 | WNAR 2023 – "Network Hawkes process models for exploring latent hierarchy in social animal interactions" |
| May 2023 | Harvard Conference in Celebration of Herman Chernoff's 100th Birthday |
| November 2022 | Data Science Day, Temple University |
| August 2021 | JSM 2021 – "Words and Insights via Text Analysis" |
| December 2020 | ICSA Applied Statistics Symposium 2020 – "Leadership in Statistics and Data Science" |
| August 2020 | JSM 2020 – " *Data for Good* at Columbia's Data Science Institute" |
| August 2019 | JSM 2019 – "Experiential Learning for Statistics Students" |
| Auguest 2018 | JSM 2018 – ""Foundation or backdrop? - The Role of Statisticians in Academic Data Science Initiatives" |
| June 2018 | ICSA Applied Statistics Symposium 2018 |
| May 2018 | The 1st Forum on Frontiers of Science and Engineering: Everything towards AI, Seattle, WA, USA. |
| May 2018 | Symposium on Data Science & Statistics (Reston, Virginia) |
| July 2017 | Program on Quantitative Methods for Drug Discovery and Development, National Singapore University |
| June 2017 | The third annual flagship conference on the Theory of Big Data, University College London, UK. |
| February 2017 | Women in Data Science (WiDS) 2017 |
| August 2016 | JSM 2016 "Phase-aligned spectral filtering for spatiotemporal dynamics" |
| July 2016 | Third Taihu International Statistics Forum (Shanghai) 'Phase-aligned spectral filtering for spatiotemporal dynamics" |
| June 2016 | Workshop on Big Data Learning for Prediction, National Sun Yat-sen University, Taiwan |
| June 2016 | Workshop on Big Data Learning for Prediction, National Tsinghua University, Taiwan. |
| June 2016 | IMS-APRM 2016 "Adaptive sparse non-linear metric learning via boosting" |
| June 2016 | ASA SLDM annual meeting "Adaptive sparse non-linear metric learning via boosting" |
| April 2016 | Columbia Data Science day. |
| Augest 2015 | JSM 2015 Discussion "The fifth V for big data" |

| | |
|---|---|
| July 2015 | "Latent Space Model for Aggregated Relational Data," 2015 Workshop on Network Data Analysis at Northeast Normal University, Changchun, Jilin, China. |
| June 2015 | "Latent Space Model for Aggregated Relational Data," Tsinghua workshop on Modern Statistics, 2015, Beijing, China. |
| Nov 2013 | "Estimation of Exponential Random Graph Models for Large Social Networks via Graph Limits" DIMACS Workshop on Statistical Analysis of Network Dynamics and Interactions, New Brunswick, NJ. |
| Oct 2013 | "A general framework of association measures, with applications to genetics." The 2013 Rao Prize Conference, Penn State University, University Park, PA. |
| August 2013 | Discussion on "social network analysis." ISI World Statistical Congress 2013, Hong Kong |
| July 2013 | "Bayesian Hierarchical Graph-Structured Model with Application to Pathway Analysis Using Gene Expression Data", IMS-China 2013, Chengdu, China. |
| August 2012 | "Social network analysis through randomly sampled respondents." SAMSI computational advertising workshop. |
| April 2012 | "Bayesian Hierarchical Graph-Structured Model with Application to Pathway Analysis Using Gene Expression Data" ENAR 2012, Washington, DC. |
| March 2011 | "Statistical methods for studying social networks using aggregated relational data". ENAR 2011, Miami, FL. |
| October 2010 | "Statistical methods for studying social networks using aggregated relational data". SAMSI complex networks modeling workshop. |
| August 2010 | "Discovering Influential Variables: A Partition-Based Learning Method to Identify Susceptible Genetic Risk Factors in Common Human Disorders," Joint Statistical Meeting 2010, Vancouver, BC, Canada. |
| July 2010 | "Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm," NCTS Workshop on "Statistics and Probability with Applications for Biology", Hsin-Chu, Taiwan. |
| June 2010 | "Studying co-regulation and inter-regulation of genes via eQTL mapping," ICSA Applied Statistics Symposium, Indianapolis, IN. |
| May 2010 | Workshop on "Statistical Issues in Analyzing Data from Diverse Sources", CCICADA (a Department of Homeland Security Center of Excellence), New Brunswick, NJ. |
| April 2010 | "Latent space model for aggregated relational data," New England Statistics Symposium, Harvard University, Cambridge, MA. |
| March 2010 | "Latent space model for aggregated relational data in the study of high-risk population for HIV+ and AIDS." ENAR 2010, New Orleans, LA. |
| August 2009 | (**Poster award**) "Latent Structural Models for Aggregated Relational Data", Joint Statistical Meeting 2009, Washington, DC. |
| July 2009 | "Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm," IMS-APRM 1st meeting, Seoul, Korea. |
| June 2009 | "Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm, Workshop on Detecting Influential Variables in High-Dimensional Data, Taida Institute of Mathematical Science, National Taiwan University, Taipei, Taiwan. |

| | |
|---|---|
| Nov 2008 | (**Peer reviewed**) "Immunohistochemical evaluation of uteri from female rhesus monkeys (Macaca Mulatta) after cyclic intramuscular estrogen treatment," American Society of Reproductive Medicine 64th Annual Meeting. (As co-author) |
| June 2008 | "Feature selection and classification based on k-nearest-neighbor patterns," Beijing International Conference on Machine Learning and Data Mining. |
| June 2008 | "Constructing Gene Association Networks for Complex Human Disorders Using the BGTA Algorithm," ICSA 2008 applied statistics symposium. |
| Oct 2007 | "Studying co-regulation and inter-regulation of genes via eQTL mapping," IEEE 2007 International Conference on Bioinformatics and BioEngineering. |
| Oct 2007 | "Evaluating the repeatability of two studies of a large number of objects: modified Kendall rank-order association test," Current and Future Trends in Nonparametrics. |
| August 2007 | "Studying co-regulation and inter-regulation of genes via eQTL mapping," Joint Statistical Meeting 2007 |
| June 2007 | (**Peer-reviewed**) "Discriminant analysis using multigene expression profiles in classification of breast Cancer," BioComp'07. |
| June 2007 | "Studying co-regulation and inter-regulation of genes via eQTL mapping," NSF Sponsored International Conference on Bioinformatics, Hangzhou, China. |
| August 2006 | Discussion on "Statistical models for networks," Joint Statistical Meeting 2006. |
| August 2006 | "Design and analysis of 'how many X' surveys," Joint Statistical Meeting 2006. |
| June 2006 | "Evaluating the repeatability of two studies of a large number of objects: modified Kendall rank-order association test," ICSA 2006 applied statistics symposium. |
| August 2005 | "A nonparametric multipoint screening method for QTL mapping," Joint Statistical Meeting 2005. |
| June 2005 | "A nonparametric multipoint screening method for QTL mapping," International Conference on Statistics in Honour of Professor Kai-Tai Fang's 65th Birthday. |
| March 2004 | "Information-driven marker selection for large scale genomic studies on complex traits," International Biometric Society, Eastern North American Region (ENAR), Spring Meeting 2004. |

---

## Teaching and Mentoring

**— Courses taught**

| | |
|---|---|
| STAT W1101 | Introduction to statistics—an introductory level course (average enroll.: 120 students). Teaching innovation—**S**tatistics **P**rojected website: http://www.stat.columbia.edu/∼tzheng/teaching/ProjectSave/.<br>Fall 2020: special "Election 2020" edition. Fall 2026: special focus on Statistical Thinking for AI. |
| STAT UN1001 | Introduction to Statistics (Fall 2026: Statistical Thinking for AI focus). |
| STAT W4315/3315 | Linear Regression Models—core requirement for M.A. in statistics program or undergraduate majors (average enroll.: 65 students). |
| STAT W4335 | Sample Surveys (average enroll.: 45 students) |

| | |
|---|---|
| STAT W8325 | Topics in Advanced Statistics—topics including relational data; network analysis; modern statistical learning and computing, with applications to biology, a topic course for PHD students (average enroll.: 10 students). |
| STAT G6101 | Statistical Modeling/Data Analysis I (Core curriculum for PHD program; average enroll.: 25 students) |
| STAT G6102 | Statistical Modeling/Data Analysis II (Core curriculum for PHD program; average enroll.: 20 students) |
| STAT G6103 | Statistical Modeling/Data Analysis III (Core curriculum for PHD program, average enroll.: 15 students). |
| STAT/CS W4242 | Introduction to Data Science. (Re-structured and designed a new course on data science for students in non-quantitative degree programs.) |
| STAT GU5243 | Applied Data Science. (Project-based learning course in data science for statistics students, average enroll.: 60 students). |
| STAT GR 6105 | Statistical Consulting and Collaborative Data Science. (Graduate students in Statistics provide free consulting services to the Columbia community while learning important collaborative skills. Average enroll.: 10 students). |

**— Short courses taught**

| | |
|---|---|
| 2026 | Short course, STAI-X 2026 (Cambridge, MA): "Agentic AI: From Zero to Infinity." |

**— Doctoral Theses supervised**

| | |
|---|---|
| In progress | Yi Han (4th year), "On Distribution-Informed Learning for Dynamical Systems." Department of Statistics, Columbia University |
| On Leave | Maria-Cristiana Girjau (5th year), Department of Statistics, Columbia University |
| 2026 | Zhewen Hou, Department of Statistics, Columbia University |
| 2024 | Diane Lu, "On Modeling Spatial Time-to-Event Data with Missing Censoring Type." Department of Statistics, Columbia University |
| 2024 | Gan Yuan, "Topics on Machine Learning under Imperfect Supervision." Department of Statistics, Columbia University (Jointly with Professor Samory Kpotufe) |
| 2023 | Reed Palmer, "Methods for Structured Data: Discrete Time Series and Networks." Department of Statistics, Columbia University |
| 2022 | Chengliang Tang, "Advances in Machine Learning for Complex Structured Functional Data" Department of Statistics, Columbia University |
| 2022 | Owen Ward, "Latent Variable Models for Events on Social Networks", Department of Statistics, Columbia University |
| 2022 | Andrew Davison, "Statistical Perspectives on Modern Network Embedding Methods" Department of Statistics, Columbia University |
| 2022 | Ashely Datta, Department of Statistics, Columbia University (Jointly with Professor Bodhi Sen) |
| 2019 | Tim Jones, "Scalable Community Detection in Massive Networks using Aggregated Relational Data," Department of Statistics, Columbia University |

2019 Jing Wu, "Point Process Models for Heterogeneous Event Time Data," Department of Statistics, Columbia University

2017 Swupnil Sahai, Department of Statistics, Columbia University
(Jointly with Professor Andrew Gelman)

2016 Lu Meng, "*Spectral Filtering for Spatio-temporal Dynamics and Multivariate Forecasts*," Department of Statistics, Columbia University

2016 Yuting Ma, "*Flexible Sparse Learning of Feature Subspaces*," Department of Statistics, Columbia University

2015 Ran He, "*A Graphon-based Framework for Modeling Large Networks*," Department of Statistics, Columbia University

2014 Hui Zhou, "*Graph structured inference for high-throughput genomic data*," Department of Biostatistics, Columbia University
(Jointly with Professor Shuang Wang)

2013 Ying Liu, "*Kernel-based association measures*," Department of Statistics, Columbia University

2012 Pengfei Zang, "*Sparse Vector Autoregressive Model for high dimensional time series*," Department of Statistics, Columbia University
(Jointly with Professor Richard Davis)

2011 Tyler McCormick, "*Statistical methods for aggregated relational data*," Department of Statistics, Columbia University

2008 Yuejing Ding, "*On feature selection and classification in high dimensions*," Department of Statistics, Columbia University

2004 Xin Yan, "*Discriminant analysis using multi-gene profiles in molecular classification of breast cancer*," Department of Statistics, Columbia University
(jointly with Professor Shaw-Hwa Lo)

2004 Hui Wang, "*A new approach of detecting influential markers for complex phenotypes with genotype data*," Department of Statistics, Columbia University
(jointly with Professor Shaw-Hwa Lo)

— **Postdoctoral researchers mentored.**

2023–2026 Shawn Li, Data Science Institute and NSF STC LEAP, Columbia University

2022-2024 Candace Agonafir, NSF STC LEAP, Columbia University

2019-2021 Isabelle Zaugg, Data Science Institute, Columbia University

2017-2019 Anna Smith, Department of Statistics, Columbia University

2016-2018 Mohammad Khabbazian, Department of Statistics, Columbia University

2015-2016 Shirin Golchi, Department of Statistics, Columbia University

2014-2015 Linda Tan, Department of Statistics, Columbia University

— **Other mentoring**

| | |
|---|---|
| 2002-present | Mentored master students, undergraduate and high school students on research projects. (2009-2010) Adeline Lo, (2010-2011) Eric Feder, Alexander Lo, (2011-2012) Seung Ah Choi, Sharon Zhu (2013-2014) Lin Yu, (2014-2015) Yuxin Zhang, (2015-2016) Michael Xie, (2017-2018) (2018-2019) Yiran Jiang (2019-2020) Dingyi Fang, Helen Jin, Lihao Xiao, (2020-2021) Erica Choi, Ribhav Talwar, Sally (Qian) Sun, Ellen Ren, Haley Elliott, Alice Choi, Antony Rossi, Aylin Dilan Demiralp, Baha Topbas, Carlyle Morgan, David Jung, Eli Hays, Ethan Malzberg, Hart Rapaport, Isaac Horwitz, Josephine Choe, Kerem Tuncer, Olivia Bobrownicki, Teddy Brodsky, Yao Lin, Yang Meng. (2021-2022) Erica Choi, Carlyle Morgan, Jiaxu Li, Serena Yuan, Amber Zhang. (2022-2023) Erica Choi, Carlyle Morgan, Fatoumata Camara. (2023-2024) Fatoumata Camara, Simpson Xu, Subashree Venkatasubramanian, Wenhe Chen, Xinyi Ke, Wenxue Li, Yue Zhao, Danning Ma, Weidong Xu, Mingzhen Zhang. (2024-2025) Subashree Venkatasubramania, Xinyi Ke, Sammy Agrawal, Luying Shu, Tianlei Zhu, Xingyao Li, Jingyu Hu, Ahmed Jaber. (2025-2026) Ahmed Jaber, Ziming Zhang, Yiyang Xiong, Hong Ru Chan, Wangshu Zhu, Zhengze Zhang, Ayon Roy, Yangyang Wang, Alex Zhang, Henry Cattermole, Kai Tai Hsu, Yilin Long, Xinai Lu, Chenglu Xing. (2026-2027) Ray Tang, Abhinav Avvaru, Ayon Roy, Ahmed Jaber, Yi Han, Yaxuan Zheng, Kartik Kumar Gounder, Khushi Pankaj Panjabi, Batuhan Duru Yeltekin, Arjun Balaji, Henry Cattermole (continuing). |
| 2021-2023 | Mentor of Diahmin Hawkins, a Bridge to PhD in STEM scholar,. |
| 2015-2020 | As faculty advisor of the Columbia Statistics Club, helped organizing hackathons with industry sponsors. |
| 2018-2019 | Organized and taught "Data for Good" bootcamp to DSI Scholars (student research interns) and Obama Fundation Scholars |
| 2016-2019 | Mentored postdoctoral fellows in the Columbia Collaboratory Program for designing data science bootcamp. |
| 2015-2017 | Mentored student instructors who taught a data science course in the Columbia Science Honor Program for high school students. |

---

## Consulting projects

| | |
|---|---|
| 2025-prenset | Consultant to Dr. Brent Stockwell, Department of Biology, Columbia University |
| 2020-present | Consultant to Dr. Nabila El-Bassel, Social Intervention Group (SIG), Columbia University School of Social Work |
| 2011-present | Consultant to Dr. Cory Abate-Shen, HICCC, Columbia University Medical Center |
| 2014-2017 | Consultant to Drs. Edward Gelmann and Shan Zha, HICCC, Columbia University Medical Center |
| 2013-2017 | Consultant to Dr. Kevin Kalinsky, HICCC, Columbia University Medical Center |
| 2011-2013 | Consultant to Bureau of Epidemiology Services, New York City Department of Health and Mental Hygiene |
| 2010 | Consultant to Drs. Alex Carballo-Dièguez and Ivan Balan at HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute and Columbia University. |
| May 2008 | Consultant to Dr. Jian Tang at Mount Sinai School of Medicine. |

| | |
|---|---|
| 2007–2008 | Consultant to Dr. Andrew Joe at the Department of Medicine, The Medical Center of Columbia University on Microarray data analysis. |
| 2005–2006 | Consultant to Dr. Doris Chang at the Department of Psychology, New School University on "Chinese immigrant women: domestic abuse and help seeking." |
| Summer 2004 | Consultant to Dr. Kwangwon Lee at the Department of Plant Pathology, Cornell University on "Characterization of the roles of light in plant-microbe interactions." |
| Summer 2003 | Consultant to Dr. Steven Schinke of the School of Social Work, Columbia University on clinical trials evaluating the effects of intervention approaches on alcohol use among adolescents. |
| Summer 2003 | Consultant to Intersystems Inc on the Prevention Program Outcomes Monitoring System survey evaluating science-based prevention programs identified and disseminated by the Center for Substance Abuse Prevention (CSAP) of the Substance Abuse and Mental Health Services Administration (SAMHSA). |
| 2002–2003 | Consultant to International Center for Health Outcomes and Innovation Research (InCHOIR), Columbia University on "Randomized Evaluation of Mechanical Assistance for Treatment of Congestive Heart Failure (REMATCH)." |

---

## Professional service

| | |
|---|---|
| 2024-present | Founding Chair (2025-), Stats Up AI, Interest Group of American Statistical Association. |
| 2026–present | Member, ASA Accreditation Task Force |
| 2026–present | ASA alternate representative to the CSAB Board |
| 2024-present | Member, NAS Committee on Frontiers of Statistics. |
| 2026–present | Member, National Academies of Sciences, Engineering, and Medicine (NASEM) Committee on Applied and Theoretical Statistics (CATS), DEPS-BMSA-18-P-84 |
| 2021-present | Member, U.S. National Committee for Mathematics (USNC/M) |
| 2023-2024 | Special Advisor to the President, Mohamed bin Zayed University of Artificial Intelligence |
| 2021-2024 | Member, Committee of Representatives to AAAS, American Statistical Association |
| 2021-2022 | Member, Nomination Committee, Institute of Mathematical Statistics |
| 2018-2020 | Chair-elect (2018), Chair (2019), Past-Chair (2020), Section on Statistical Learning and Data Science, American Statistical Association |
| 2016-2019 | Advisory board member, Stats.org |
| 2014-2017 | Board of directors, International Chinese Statistical Association (ICSA) |

### — Editorial and review service

| | |
|---|---|
| 2020-present | Member, Executive Board, Springer Series on "Teaching Data Science" |
| 2013-present | Head of Faculty (2020-), Faculty member (2013-2020), Faculty Opinions (formerly, *F1000 Prime*). |
| 2019-2024 | Associate editor, Science Advances |

| | |
|---|---|
| 2018-2024 | Reviewer, AMS/Mathematical Reviews |
| 2020 | Panel member, National Science Foundation |
| 2019 | Panel member, National Science Foundation |
| 2013-2019 | Associate editor, Statistical Analysis and Data Mining (SAM). |
| 2017 | Panel member, National Science Foundation |
| 2016 | Panel member, National Science Foundation |
| 2015 | Panel member, National Science Foundation |
| 2013-2014 | ASA Forensic Science Advisory Committee (FSAC) GAO group |
| 2011-2014 | Review editor, Frontiers in Genetic Architecture—a specialty of Frontiers in Genetics. |
| 2007-2013 | Associate editor, Journal of the American Statistical Association, Applications and Case Studies. |
| 2011-2012 | Guest editor, for a special issue on networks, Statistics and its Interface. |
| 2011-2012 | JASA ACS editor search committee. |

— **Referee**

| ASA SLDM student paper competition 2015, 2016, 2017 | BioComp'07 | Behavior Genetics | Bioinformatics | Biometrika | BMC Genomics | Communications in Statistics | Human Heredity | IEEE BIBE 07 | IEEE transactions on signal processing | Journal of Official Statistics | Journal of Statistical Planning and Inference | Journal of the American Statistical Association | National Security Agency | Nature | New Media and Society | NSF | PLoS genetics | Sociological Methodology | PNAS | Transactions on Big Data | Statistical Applications in Genetics and Molecular Biology | Statistical Methodology | Statistics in Bioscience | Prentice Hall | The American Statistician | U.S. Army Research | Wiley | American Medical Informatics Association Symposium 2005 |

— **Conference/invited sessions organizers and chairs**

| | |
|---|---|
| Jan 2027 | Organizer, Stats Up AI AI Leadership Academy (AILA) |
| 2026 | General Co-Chair, AASF AIX Summit \| East 2026 |
| 2024 | Co-chair, MBZUAI's "AI Quorum on Statistics for the Future of AI." |
| 2024 | Organizer, Invited Session WNAR "Statistical Models for Population Disease Dynamics" |
| 2024 | Co-organizer, AAAS "Changing Ice Sheets and Glaciers: Understanding Data, Models, and Impacts" |
| 2023 | Organizer, JSM invited session "Advances in Statistical Learning for Complex Networks." |
| 2023 | Organizer, JSM Late-Breaking Session on "It Takes More Than a Village: Capacity-Building Efforts for the Statistical Community." |
| 2020 | Member, Program Committee, ACM-IMS Foundations of Data Science 2020 |
| 2019-2020 | General co-Chair, ASA SLDS biannual conference 2020 in Irvine, CA. |
| 2019 | Member, StatFest 2019 Planning Committee. |

| | |
|---|---|
| 2019 | Publicity co-chair, IEEE DSAA 2019 |
| 2019 | Member, Program Committee, NESS 2019 |
| 2017-2018 | Chair, local committee for ASA SLDS biannual conference 2018 in New York, NY. |
| 2018 | Organized invited session on "Statistical Challenges in Large-Scale Data Mining", Symposium on Data Science & Statistics. |
| 2018 | Organized invited panel on "Foundation or backdrop? - The Role of Statisticians in Academic Data Science Initiatives", Joint Statistical Meeting, 2018. |
| 2017-2018 | Member, committee on student paper competition for JSM 2018, section on Statistical Learning and Data Mining, ASA. |
| 2016-2017 | Chair, committee on student paper competition for JSM 2017, section on Statistical Learning and Data Mining, ASA. |
| 2016 | Organized special session on "Statistical Learning for Data Science" in ASA sponsored 2016 IEEE Conference on Data Science and Applied Analytics. |
| 2015-2016 | Chair, committee on student paper competition for JSM 2016, section on Statistical Learning and Data Mining, ASA. |
| 2015 | Organized topic-contributed sessions for JSM 2015: "Statistical methods for observational health studies," and "Scalable Bayesian learning and computing." |
| 2014-2015 | Co-chair, committee on student paper competition for JSM 2015, section on Statistical Learning and Data Mining, ASA. |
| 2012-2014 | Organizing committee, member, SAMSI 2013-2014 program on Computational Methods in Social Sciences. |
| 2011-2013 | Scientific program committee vice-chair, ISI World Statistics Congress 2013, Hong Kong. |
| 2009-2011 | Executive committee, Program Chair, ICSA Applied Statistics Symposium 2011, New York |
| 2010 | Organized invited session for JSM 2010 "Latent space models for network analysis" |
| 2010 | Organized IMS sponsored invited session for ENAR 2010 "Studying genetic and environmental risk factors of complex human disorders and their interactions" |
| 2007 | Program committee, IEEE 2007 International Conference on Bioinformatics and BioEngineering. |
| 2007 | Organized invited session JSM 2007: *Statistical methods for gene regulatory networks.* |
| 2006–present | Organized invited sessions for ICSA applied statistics symposium 2006: *Recent contributions to nonparametric methodologies in biomedical research;* 2008: *Designs and analyses of genome-wide association studies.* |
| 2005 | Organized topic contributed session JSM 2005: *Recent statistical contributions to Genetics and Epidemiology.* |
| 2003–present | Chaired sessions for Joint Statistical Meetings: 2003, 2005 and 2007. |

— **University/Departmental service**

| | |
|---|---|
| 2019-present | Member, Faculty Advisory Board, Columbia Beijing Global Center |
| 2020-present | Member, Faculty Committee, Science Of Learning Research (SOLER) Initiative |
| 2020-present | Member, Provost's Faculty Committee on Educational Innovation |
| 2026 | Member, Columbia Presidential Task Force on Constructive Dialogue |
| 2026–present | Faculty lead, aiX Faculty Fellowship on AI education, Columbia University |
| 2014-present | Member, HPC (Hotfoot/Yeti/Habenero/Terramotto) OC Executive committee |
| 2014-2025 | Member, Shared Research Computing Policy Advisory Committee (SRCPAC), Columbia University |
| 2015-present | Faculty advisor, Columbia Statistics Club |
| 2026–2027 | Member, M.A. Advisory Committee, Department of Statistics, Columbia University |
| 2026–2027 | Member, Senior Faculty Search Committee, Department of Statistics, Columbia University |
| 2026–2027 | Member, Ph.D. Admissions Committee, Department of Statistics, Columbia University |
| 2013-2020 | Chair (2017 -2020 ) and member, Education committee, Data Science Institute, Columbia University |
| 2017-2020 | Associate Director for Education, Data Science Institute, Columbia University |
| 2017-2020 | Member, Research Subcommittee of the Provost's Advisory Committee on the Libraries, Columbia University |
| 2019 | Chair, cross-department group on computational and theoretical sciences, Columbia A&S Science Plan |
| 2017-2019 | Organizer, "Data for Good" public talk series, Data Science Institute, Columbia University |
| 2018-2019 | Member, Foundation for Research Computing, Columbia University |
| 2017-2019 | Member, Academic Review Committee (ARC), Arts & Sciences, Columbia University |
| 2012-2018 | Chair, Center on Foundations of Data Science, Data Science Institute, Columbia University |
| 2018 | Member, methodology tenure track faculty searching committee, Department of Statistics, Columbia University |
| 2018 | Member, hiring committee, Department of Statistics, Columbia University |
| 2018 | Member, lecturer review committee, school of professional studies, Columbia University |
| 2015-2017 | Member, data and society task force, Columbia University |
| 2017 | Department representative, Equality committee, Arts & Sciences, Columbia University |
| 2017 | Member, EPPC subcommittee on Coordination of Instructional Staffing and Scheduling, Arts & Sciences, Columbia University |
| 2017 | Chair, faculty search committee for open-rank position, Department of Statistics, Columbia University |

| | |
|---|---|
| 2016-2017 | Member, hiring committee for lecturer in discpline, Department of Statistics, Columbia University |
| 2016 | Chair, hiring committee for term assistant professors, Department of Statistics, Columbia University |
| 2015-2016 | Series creator, ColumbiaX series on "data science and analytics" on edX. |
| 2015 | Lenfest junior faculty grant review committee, Columbia University |
| 2015 | Review committee, ROADS grant, Data Science Institute, Columbia University |
| 2013-2014 | Member, hiring committee for lecturer of Analytics, School of Continuing Education, Columbia University |
| 2007-2008 | Director of graduate study (Ph.D. program), Department of Statistics, Columbia University. |
| 2006-07, 2002-03 | Seminar chair, Department of Statistics, Columbia University. |
| 2004–present | Doctoral students mentoring: Yuejing Ding (2004), Jiayi Chen (2005), Xiaoru Wu (2006), Li Song (2006), Tyler McCormick (2007), Ying Liu (2009), Lu Meng (2011), Yuanjun Gao (2012), Shuaiwen Wang (2013), Yixin Wang (2014), Chengliang Tang (2016), Yi Han (2023), Joule Volez (2024), Department of Statistics, Columbia University. |
| 2002–present | Computing committee, committee on curriculum, committee on junior search, committee on PhD admissions, and other ad hoc committees, Department of Statistics, Columbia University. |

— **Doctoral Defense Committees**

| | |
|---|---|
| 2025 | Casey Bradshaw, Department of Statistics, Columbia University |
| 2024 | (Oral Exam) Zhewen Hou, Department of Statistics, Columbia University |
| | Diane Lu, Department of Statistics, Columbia University |
| | Gan Yuan, Department of Statistics, Columbia University |
| | Julien Boussard, Department of Statistics, Columbia University |
| 2023 | Ashley Datta, Department of Statistics, Columbia University |
| | (Oral Exam) Julien Boussard, Department of Statistics, Columbia University |
| | Diane Lu, Department of Statistics, Columbia University |
| | Baihan Lin, Department of Systems Biology, Columbia University |
| | Reed Palmer, Department of Statistics, Columbia University |
| | Nicholas Galbraith, Department of Statistics, Columbia University |
| 2022 | Andrew Davison, Department of Statistics, Columbia University |
| | Owen Ward, Department of Statistics, Columbia University |
| | David Thibodeaux, Department of Biomedical Engineering, Columbia University |
| | Bridget, Ratcliffe, Department of Statistics, Columbia University |
| | Chengliang Tang, Department of Statistics, Columbia University |

| | |
|---|---|
| | Ashley Datta, Department of Statistics, Columbia University |
| | (Oral Exam) Ari Blau, Department of Statistics, Columbia University |
| 2021 | (Oral Exam) Reed Palmer, Department of Statistics, Columbia University |
| | (Oral Exam) Ashley Datta, Department of Statistics, Columbia University |
| | Ding Zhou, Department of Statistics, Columbia University |
| | Yuling Yao, Department of Statistics, Columbia University |
| 2020 | (Oral Exam) Bridget Lynn Ratcliffe, Department of Statistics, Columbia University |
| | (Oral Exam) Owen Ward, Department of Statistics, Columbia University |
| | (Oral Exam) Andrew Davison, Department of Statistics, Columbia University |
| | (Oral Exam) David Nicholas Thibodeaux, Department of Biomedial Engineering, Columbia University |
| | (Oral Exam) Baihan Lin, Systems Biology & Computational Neuroscience, Columbia University |
| | Lucas, Gloege, Department of Earth and Environmental Sciences, Columbia University |
| | Sihan Huang, Department of Statistics, Columbia University |
| | Kaiyi Zhu, Department of Electrical Engineering, Columbia University |
| | Adji Bousso Dieng, Department of Statistics, Columbia University |
| | Yayun (Lydia) Hsu, Department of Statistics, Columbia University |
| 2019 | Tim Jones, Department of Statistics, Columbia University |
| | Jing Wu, Department of Statistics, Columbia University |
| | (Oral Exam) Chengliang Tang, Department of Statistics, Columbia University |
| | (Oral Exam) Ding Zhou, Department of Statistics, Columbia University |
| | (Oral Exam) Yuling Yao, Department of Statistics, Columbia University |
| | (Oral Exam) Baihan Lin, Integrated in CMBS/C2B2, Columbia University |
| 2018 | Daniel Thomas Bartilson, Department of Civil Engineering, Columbia University |
| | Jihui Lee, Department of Biostatistics, Columbia University |
| | Feihan Lu, Department of Statistics, Columbia University |
| | Ying Ma, Department of Biomedical Engineering, Columbia Univerisity |
| | (Oral Exam) Sihan Huang, Department of Statistics, Columbia University |
| | (Oral Exam) Lydia Hsu, Department of Statistics, Columbia University |
| 2017 | Swupnil Sahai, Department of Statistics, Columbia University |
| | Ruma Basu, Department of Statistics, Columbia University |
| | Haolei Weng, Department of Statistics, Columbia University |
| | Yuanjun Gao, Department of Statistics, Columbia University |

(Oral exam) Timothy Jones, Department of Statistics, Columbia University

2016 (Oral Exam) Feihan Lu, Department of Statistics, Columbia University

(Oral Exam) Jihui Lee, Department of Biostatistics, Columbia University

(Oral Exam) Ying Ma, Department of Biomedical Engineering, Columbia Univerisity

Yuting Ma, Department of Statistics, Columbia University

Adeline Lo, Department of Sociology, University of California at San Diego

Lu Meng, Department of Statistics, Columbia University

Diego Franco Saldana, Department of Statistics, Columbia University

(Oral Exam) Jing Wu, Department of Statistics, Columbia University

2015 Michael Agne, Department of Statistics, Columbia University

(Oral Exam) Haolei Weng, Department of Statistics, Columbia University

Ran He, Department of Statistics, Columbia University

(Oral Exam) Adeline Lo, Department of Political Science, University of California at San Diego

(Oral Exam) Lu Meng, Department of Statistics, Columbia University

2014 (Oral Exam) Diego Franco, Department of Statistics, Columbia University

Mengqian Lu, Department of Earth and Environmental Engineering, SEAS, Columbia University

Ruixue Fan, Department of Statistics, Columbia University

Hui Zhou, Department of Biostatistics, Columbia University

(Oral Exam) Michael Agne, Department of Statistics, Columbia University

(Oral Exam) Ruixue Fan, Department of Statistics, Columbia University

(Oral Exam) Ran He, Department of Statistics, Columbia University

2013 (Oral Exam) Hui Zhou, Department of Biostatistics, Columbia University

Ying Liu, Department of Statistics, Columbia University

Chien Hsun Huang, Department of Statistics, Columbia University

Vincent Dorie, Department of Statistics, Columbia University

Bo Qian, Department of Statistics, Columbia University

2012 (Oral Exam) Ying Liu, Department of Statistics, Columbia University

(Oral Exam) Chien Hsun Huang, Department of Statistics, Columbia University

(Oral Exam) Vincent Dorie, Department of Statistics, Columbia University

Pengfei Zang, Department of Statistics, Columbia University

2011 (Oral Exam) Pengfei Zang, Department of Statistics, Columbia University

Xiaoru Wu, Department of Statistics, Columbia University

| | |
|---|---|
| | Amal Moussa, Department of Statistics, Columbia University |
| | Tyler McCormick, Department of Statistics, Columbia University |
| | (Oral Exam) Gongjun Xu, Department of Statistics, Columbia University |
| 2010 | (Oral Exam) Tyler McCormick, Department of Statistics, Columbia University |
| 2008 | (Oral Exam) Xiaoru Wu, Department of Statistics, Columbia University |
| | Yuejing Ding, Department of Statistics, Columbia University |
| | Ivan Iossifov, Department of Medical Informatics, Columbia University |
| 2006 | (Oral Exam) Yihong Zhao, Department of Biostatistics, Columbia University |
| | Iuliana Ionita, Department of Computer Science, New York University |
| | (Oral Exam) Shouhao Zhou, Department of Statistics, Columbia University |
| | Jouni Kerman, Department of Statistics, Columbia University |
| 2005 | (Oral Exam) Jouni Kerman, Department of Statistics, Columbia University |
| | Jing Han, Division of Biostatistics, New York University |
| 2004 | (Thesis outline defense) Jing Han, Division of Biostatistics, New York University |
| | Hui Wang, Department of Statistics, Columbia University |
| | Xin Yan, Department of Statistics, Columbia University |
| 2003 | Cristian Pasarica, Department of Statistics, Columbia University |
| | Zaiying Huang, Department of Statistics, Columbia University |
| 2002 | Yuhang Xing, Department of Finance, Columbia University |

---

## Memberships

American Statistical Association (Fellow)

American Association for the Advancement of Science (Fellow)

Asian American Academy of Science and Engineering (Life-time member)

Asian American Scholar Forum (Fellow)

Royal Statistical Society

Institute of Mathematical Statistics (Fellow)

International Chinese Statistical Association (Life-time member)

International Statistical Institute (elected)