BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Wing Hung Wong

eRA COMMONS USER NAME (credential, e.g., agency login): wwong1

POSITION TITLE: Professor, Department of Statistics and Department of Biomedical Data Science

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of California, Berkeley, CA	B.A.	06/1976	Mathematics
University of Wisconsin, Madison, WI	M.S.	06/1978	Statistics
University of Wisconsin, Madison, WI	M.S.	06/1982	Computer Science
University of Wisconsin, Madison, WI	Ph.D	09/1980	Statistics

A. Personal Statement

My current research interests include gene regulatory analysis, analysis of single cell genomics data, interpretation of noncoding variants and polygenic risk prediction, and disease detection from cell-free DNA analysis. These areas require new methods in statistics, machine learning, mathematical modeling and computation. In the past, I have contributed useful concepts and methods both in mainstream statistics and in bioinformatics. Examples include Monte Carlo algorithms in Bayesian computation, asymptotic inference in high or infinite dimensional problems. My group have developed statistical methods and computer programs for the analysis of bulk or single cell sequencing data and have applied these methods to the study of gene regulatory networks based on integrative modeling of multiple types of genomics data from diverse cell types. We have also used these cell type specific regulatory networks to improve the analysis of case-control data based on whole genome sequencing data. Coming from a mathematical statistics background, I am interested in working with experimental collaborators in the analysis of data from the latest genomics technologies, and with medical collaborators in the interpretation of their clinical data.

B. Positions, Scientific Appointments, and Honors

Positions

2015- Professor, Department of Biomedical Data Science, Stanford University
2012- Stephen R. Pierce Family Goldman Sachs Professor, Stanford University
2004- Professor, Department of Statistics, Stanford University
2000-2004 Professor, Department of Statistics & Department of Biostatistics, Harvard University
1997-2000 Professor, Departments of Statistics, University of California at Los Angeles
1994-1997 Professor, Chairman, Department of Statistics, Chinese University of Hong Kong.
1980-1994 Assistant, Associate to Full Professor, Department of Statistics, University of Chicago

Honors

2023 IMS Grace Wahba Award & Lecture
2021 COPSS Distinguished Achievement Award and Lectureship (formerly known as R. A. Fisher Award and Lectureship)
2019 Statistician of the Year (American Statistical Association, Chicago Chapter)
2015 Elected to the Hong Kong Academy of Sciences
2010 Elected to Academia Sinica
2009 Elected to National Academy of Sciences, USA

2006	Bahadur Lecture
2002	Neyman Lecture
2002	Fellow of American Association for the Advancement of Sciences
1998	Fellow of American Statistical Association
1993	COPSS Presidents' Award
1991	Fellow of Institute of Mathematical Statistics
1987	Guggenheim Fellow

C. Contributions to Science

I) Bayesian Computation

Sampling-based methods for Bayesian inference such as Markov Chain Monte Carlo (MCMC) or sequential importance sampling are among the most widely used statistical inference methods. I have made important contributions in these areas. We proposed the data augmentation algorithm (Tanner and Wong 1987) which was an iterative sampling algorithm for drawing parameter values from posterior distributions. This work had significantly influenced subsequent research such as Gelfand & Smith's paper in 1990 that ushered in the use of MCMC in general Bayesian inference. Later, working with various collaborators, I had also contributed to the development of sequential importance sampling (Kong, Liu, Wong 1994), collapsed Gibbs sampler, evolutionary Monte Carlo, equi-energy sampling (Kou, Zhou, Wong 2006), and Mini-batch MCMC methods. Our recent work on mini-batching established its asymptotic equivalence to tempering (Wu, Wang and Wong 2020) and is relevant to current AI applications as mini-batching is necessary for learning from very large data sets.

1) Martin Tanner and Wing Hung Wong (1987) Calculation of posterior distributions by data augmentation. (Discussion paper). *Journal of the American Statistical Association*, 82, 528--540
2) Augustine Kong, Jun Liu and Wing Hung Wong (1994) Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89, 278--288
3) Sam Kou, Qing Zhou and Wing Hong Wong (2006) Equi-energy sampling and its application to statistical inference and statistical mechanics. (Discussion paper) *Annals of Statistics*, 34:1581-1619.
4) Tung-Yu Wu, Y. X. Rachel Wang & Wing H. Wong (2020) Mini-Batch Metropolis–Hastings With Reversible SGLD Proposal, *Journal of the American Statistical Association* DOI: 10.1080/01621459.2020.1782222

II) Statistical Theory and Methods

Modern statistical models often involve infinite dimensional parameters such as a hazard function in a Cox Model, or a density function for an i.i.d. sample. I have made substantial contributions to these nonparametric and semi-parametric models. Wong (1986) developed asymptotic properties of "partial likelihood". Later, Xiaotong Shen and I discovered that sieve-MLE (i.e. MLE obtained by maximizing over an approximation of the parameter space with reduced complexity) can attain the optimal rate of convergence in situations when the classical MLE converges at a slow rate (Shen and Wong 1994), and established a fundamental probability inequality bounding the maximum of the likelihood surface outside of a Hellinger ball around the true density (Wong & Shen 1995). This result is important for later development such as the analysis of Shen and Wasserman (2001) on the concentration rate of posterior distributions. More recently, we proved the asymptotic consistency of Hinton's contrastive divergence algorithm which has been a long standing open problem.

1) Xiaotong Shen and Wing Hung Wong (1994) Convergence rate of sieve estimates. *Annals of Statistics*, 22, 580--615.
2) Wing Hung Wong and Xiaotong Shen (1995) Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Annals of Statistics*, 23, 339--362
3) Tung-Yu Wu, Bai Jiang, Yifan Jin and Wing H. Wong (2018) Convergence of contrastive divergence algorithm in exponential family. *Annals of Statistics*, Volume 46, Number 6A (2018), 3067-3098. arXiv:1603.05729
4) Qiao Liu, Zhongren Chen, Wing Hung Wong (2024) An encoding generative modeling approach to dimension reduction and covariate adjustment in causal inference with observational studies Proc Natl Acad Sci U S A, June 4, 2024, 121 (23) e2322376121

III) Analysis of microarray and RNA-seq data

We have leveraged our statistical expertise to develop methods and software for the analysis of gene expression microarrays, RNA-seq and ChIP-seq data. We introduced the model-based approach for the analysis of gene expression by oligonucleotide arrays (Li and Wong, 2001) and implemented it in the widely used software dChip. This work had significant influence on other important model-based analysis tools such as RMA (Irizarry et al 2003) and MACS (Zhang et al 2008). Later, when NGS based methods such as RNA-seq and ChIP-seq were being developed, we are among the first to provide statistical tools for analyzing the resulting data. For example, Jiang and Wong (2009) proposed the first statistical model for isoform specific gene expression from RNA-seq. Au et al (2013) developed a hybrid 2nd generation and 3rd generation sequencing approach for isoform discovery and applied the methods to discover many new transcripts in human embryonic stem cell.

1) Hongkai Ji, Hui Jiang, Wenxiu Ma, David S Johnson, Richard M Myers, Wing H Wong (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nature Biotechnology*. 26, 1293-1300, doi:10.1038/nbt.1505
2) Hui Jiang and Wing Hung Wong (2009) Statistical Inferences for isoform expression in RNA-seq. *Bioinformatics* 25(8):1026-1032; doi:10.1093/bioinformatics/btp113
3) Kin Fai Au, Vittorio Sebastiano, Pegah Tootoonchi Afshar, Jens Durruthy Durruthy, Lawrence Lee, Brian A. Williams, Harm van Bakel, Eric E. Schadt, Renee A. Reijo-Pera, Jason G. Underwood, and Wing Hung Wong (2013) Characterization of the human ESC transcriptome by hybrid sequencing *Proceedings of the National Academy of Sciences* 2013 110 (50) E4821-E4830
4) Shaowei Li, Xi Chen, Jingxue Xin, Bowen Liu, Bo Liu, Min Hu, and Wing Hung Wong (2025) Human stem cells with in vivo high plasticity generated by cell-cell communication Proc Natl Acad Sci U S A, March 11, 2025, 122 (11) e2413043122

IV) Gene regulatory systems and computational genomics

We are interested in the computational inference of gene regulatory networks based on functional genomics data. Early on, when only gene expression data was available, we used cell sorting and time course expression arrays to characterize transcriptional regulation in mESC differentiation (Zhou 2007, PNAS 104(42):16438-43). However, we found that expression data by itself was not sufficient for gene regulatory analysis, and methods such as ChIP-chip or ChIP-seq, while powerful, were regulator-specific and could not provide a comprehensive picture of the whole regulatory system. Therefore, when the technology for mapping accessible regions become available, we were among the first to produce methods for gene regulatory network analysis based on paired RNA-seq and ATAC-seq data across multiple cellular contexts (Zhana 2017). More recently, we have developed several tools for analyzing gene regulatory systems based on multiple types of single cell genomics data (Liu 2021, Y Lin 2022). Our collaborative work in this area had resulted in substantial scientific findings (X Lin 2022, Zhou 2024)

1) Qiao Liu, Shengquan Chen, Rui Jiang and Wing Hung Wong (2021) Simultaneous deep generative modeling and clustering of single cell genomic data. *Nature Machine Intelligence*, VOL 3, June 2021, pages536–544. https://doi.org/10.1038/s42256-021-00333-y
2) Yingxin Lin, Tung-Yu Wu, Sheng Wan, Jean Y.H. Yang, Wing H. Wong and Y. X. Rachel Wang (2022) scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. *Nature Biotechnology*, Published online: 20 January2022. https://doi.org/10.1038/s41587-021-01161-6
3) Bo Zhou, Joseph G. Arthur, Hanmin Guo, Taeyoung Kim, Yiling Huang, Reenal Pattni, Tao Wang, Soumya Kundu, Jay X.J. Luo, HoJoon Lee, Daniel C. Nachun, Carolin Purmann, Emma M. Monte, Annika K. Weimer, Ping-Ping Qu, Minyi Shi, Lixia Jiang, Xinqiong Yang, John F. Fullard, Jaroslav Bendl, Kiran Girdhar, Minsu Kim, Xi Chen, PsychENCODE Consortium, William J. Greenleaf, Laramie Duncan, Hanlee P. Ji, Xiang Zhu, Giltae Song, Stephen B. Montgomery, Dean Palejev, Heinrich zu Dohna, Panos Roussos, Anshul Kundaje, Joachim F. Hallmayer, Michael P. Snyder, Wing H. Wong, Alexander E. Urban (2024) Detection and analysis of complex structural variation in human genomes across populations and in brains of donors with psychiatric disorders. Cell, 2024 Nov 14;187(23):6687-6706.e25. doi: 10.1016/j.cell.2024.09.014. Epub 2024 Sep 30
4) Wanwen Zeng, Hanmin Guo, Qiao Liu and Wing Hung Wong (2025) Improving polygenic prediction from whole-genome sequencing data by leveraging predicted epigenomic features. Proc Natl Acad Sci., 2025 Jun 17;122(24):e2419202122. doi: 10.1073/pnas.2419202122. PMCID: PMC12184400