

Wenpin Hou (Ph.D.)

E3011 Bloomberg School of Public Health
Johns Hopkins University, Baltimore, MD 21210, United States

Tel: (+1) 443-635-4567
Email: whou10@jhu.edu

EDUCATION AND TRAINING

Degrees

- 9/2013-8/2017 **Ph.D., Mathematics, The University of Hong Kong** (with University Fellowships), Hong Kong
Mentor: Dr. Wai-Ki Ching
- 9/2009-7/2013 **B.Sc., Information and Computational Science, Sun Yat-sen University**, Guangzhou
Mentor: Dr. Tianshou Zhou

Postdoc Training

- 8/2019-present **Biostatistics, Johns Hopkins University**, Baltimore
Mentor: Dr. Stephanie Hicks, Dr. Hongkai Ji
Focus on developing statistical and computational methods to tackle the challenges in single-cell genomics and epigenomics using multi-omics data, such as scRNA-seq data imputation benchmark, scATAC-seq signal enhancement, differential pseudotime analysis, inferring DNA methylation and its spatial landscape, and multiple collaborative research on investigating glioblastoma, pediatric glioma, pancreatic cancer, non-small-cell lung cancer and PD-1 therapy
- 8/2017-8/2019 **Computer Science, Johns Hopkins University**, Baltimore
Mentor: Dr. Aravinda Chakravarti, Dr. Suchi Saria
Focus on developing statistical models for analyzing time-series genomic data in cell reprogramming, the reconstruction of gene regulatory networks using scATAC-seq and scRNA-seq data, and characterizing the relationship between chromatin accessibility and gene expression

Visiting Researcher Training

- 7-8/2017 **Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center**, Houston
Advisor: Dr. Anil Korkut. Work on genomics-guided discovery of effective combination therapy in cancer
- 5-8/2015 **Bioinformatics Center, Kyoto University**, Kyoto
- 2, 5-8/2016 Advisor: Dr. Tatsuya Akutsu. Work on the construction and controllability problems in Boolean Networks and Probabilistic Boolean Networks
- 11-12/2015 **Department of Mathematics, Renmin University of China**, Beijing
Collaborator: Dr. Hao Jiang. Work on developing Hadamard Kernel SVM for breast cancer outcome prediction
- 6-8/2014 **Bioinformatics Lab, Soka University**, Tokyo
Advisor: Dr. Kiyoko F. Aoki-Kinoshita. Work on the auto-reconstruction of glycosylation networks

Honorary Research Associate

- 4/2018-6/2020 **Department of Mathematics, The University of Hong Kong**, Hong Kong

HONORS & AWARDS

- 2021 NIH Pathway to Independence Award (K99/R00), NIH/NHGRI
- 2019 First-place Winning Poster at the 13th Annual Symposium and Poster Session on Genomics and Bioinformatics
- 2013-2017 University Postgraduate Fellowships
- 2013-2017 Postgraduate Scholarship
- 2016 Excellent Teaching Award
- 2016 Doris Chen Postgraduate Travel Grants
- 2009-2012 Academic Award for Excellent Students

- 2012 Vice Chancellor Xu Chongqing Award for Academic Excellence
- 2012 Honorable Mention of 2012 Mathematical Contest in Modelling, COMAP, NSA
- 2011 National Endeavor Award for Academic Excellence (ranked 3/119)
- 2011 Second Prize of 2011 Chinese Mathematical Contest in Modelling, MOE of PRC, CSIAM

RESEARCH INTERESTS

Genomics and epigenomics: develop statistical and computational methods in genomics and epigenomics; gene regulatory networks inference; mathematical and computational methods in medicine, clinics, and health care

Applied mathematics: complex networks, optimization, probability and their applications

RESEARCH GRANT

NIH/NHGRI K99HG011468: Computational Methods for Inferring Single-cell DNA Methylation and its Spatial Landscape.

Dates: 03/2021 – 03/2026. Principal Investigator: Wenpin Hou, PhD.

Responsibility: Principal Investigator (PI).

PUBLICATIONS

Journal Papers / Preprint

Single-cell genomics

1. **Hou, W.**, Ji, Z., Chen, Z., Wherry, E.J., Hicks, S.C. and Ji, H., 2021. A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. [bioRxiv](#). Package: [Lamian](#).
2. **Hou, W.** and Ji, Z., Single-Cell Unbiased Visualization With SCUBI. [SSRN 3881357](#). Package: [SCUBI](#).
3. Ji, Z., Zhou, W., **Hou, W.** and Ji, H.*, 2020. SCATE: Single-cell ATAC-seq Signal Extraction and Enhancement. [Genome Biology, 21,161 \(2020\)](#). PMID: 32620137. PMCID: PMC7333383. Packages: [SCATE](#), [SCATEData](#).
4. **Hou, W.**, Ji, Z., Ji, H.* and Hicks, S.C.*, 2020. A Systematic Evaluation of Single-cell RNA-sequencing Imputation Methods. [Genome Biology 21, 218 \(2020\)](#). PMID: 32854757. PMCID: PMC7450705.

Boolean networks

5. **Hou, W.**, Ruan, P., Ching, W.K. and Akutsu, T.*, 2019. On the number of driver nodes for controlling a Boolean network when the targets are restricted to attractors. [Journal of Theoretical Biology, 463, pp.1-11](#).
6. **Hou, W.**, Tamura, T., Ching, W.K. and Akutsu, T., 2016. Finding and analyzing the minimum set of driver nodes in control of Boolean networks. [Advances in Complex Systems, 19\(03\), p.1650006](#).

Glycosylation networks

7. **Hou, W.**, Qiu, Y., Hashimoto, N., Ching, W.K. and Aoki-Kinoshita, K.F.*, 2016. A systematic framework to derive N-glycan biosynthesis process and the automated construction of glycosylation networks. [BMC Bioinformatics, 17\(7\), p.240](#).

Epigenetics

8. Xu, R., Hong, X.*, Zhang, B., Huang, W., **Hou, W.**, Wang, G., Wang, X., Igusa, T., Liang, L. and Ji, H., 2021. DNA methylation mediates the effect of maternal smoking on offspring birthweight: a birth cohort study of multi-ethnic US mother–newborn pairs. [Clinical epigenetics, 13\(1\), pp.1-13](#). PMID: 33663600. PMCID: PMC7931602.

Psychiatry

9. Ji, Y., Azuine, R.E., Zhang, Y., **Hou, W.**, Hong, X., Wang, G., Riley, A., Pearson, C., Zuckerman, B. and Wang, X.*, 2019. Association of cord plasma biomarkers of in utero acetaminophen exposure with risk of attention-deficit/hyperactivity disorder and autism spectrum disorder in childhood. [JAMA psychiatry, 77\(2\), pp.180-189](#).. PMID: 31664451. PMCID: PMC6822099.
(featured in [NIH news](#), [Reuters health](#), [MedPage Today](#), [meaww](#), [LinksMedicus](#), [TechnologyNetworks](#))

Machine learning

10. Jiang, H.*, Qiu, Y., **Hou, W.**, Cheng, X., Yim, M.Y. and Ching, W.K., 2018. Drug Side-Effect Profiles Prediction: From Empirical to Structural Risk Minimization. [IEEE/ACM transactions on computational biology and bioinformatics, 17\(2\), pp.402-410](#).
11. Jiang, H.*, Ching, W.K., Cheung, W.S., **Hou, W.** and Yin, H., 2017. Hadamard kernel SVM with applications for breast cancer outcome predictions. [BMC Systems Biology, 11\(7\), p.138](#).
12. Jiang, H.*, Ching, W.K. and **Hou, W.**, 2016. On orthogonal feature extraction model with applications in medical prognosis. [Applied Mathematical Modelling, 40\(19-20\), pp.8766-8776](#).
13. **Hou, W.***, Chen, Y. and Zhang, Y., Investigation of Heavy Metal Pollution on Urban Topsoil, *Economic Life Digest* (in Chinese), 15 (2012) pp.204-206. [ISSN1009 – 5535]

Conference Proceedings

14. Qiu, Y., Cheng, X., **Hou, W.** and Ching, W.K.*, On classification of biological data using outlier detection, *The 12th International Symposium on Operations Research and its Applications (ISORA)*, Luoyang, China, 2015, pp.144-150.

15. Cheng, X., Ching, W.K., **Hou, W.** and Kinoshita, K.* , A multiple linear regression model for structure of N-linked oligosaccharides, *The 12th International Symposium on Operations Research and its Applications (ISORA)*, Luoyang, China, 2015, pp.151-157.
16. Cheng, X., Qiu, Y., **Hou, W.** and Ching, W.K.* , A semi-tensor product approach for probabilistic Boolean networks, *IEEE Proceedings of the 8th International Conference on Systems Biology (ISB)*, Qingdao, China, 2014, pp.88-93.

Posters

17. A systematic evaluation of single-cell RNA-seq imputation methods, *13th Annual Symposium and Poster Session on Genomics and Bioinformatics*, Johns Hopkins University, Baltimore, USA, Oct 17, 2019. (awarded as *first-place winning poster*)
18. Causal gene regulatory network construction using single-cell RNA-seq and single-cell ATAC-seq data, *11th annual RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges*, New York University, New York, USA, Dec 8-10, 2018.
19. On orthogonal feature extraction model with applications in cancer prediction, University of Cadiz, Cadiz, Spain, May 18, 2016.

Thesis

20. **Hou, W.** (2017) Mathematical modelling and optimization in biological networks and data. PhD thesis. The University of Hong Kong.

TALKS

- | | |
|---|---------|
| "A Systematic Evaluation of Single-Cell RNA-seq 2:45 Imputation Methods", ENAR 2020 Spring Meeting | 3/2020 |
| "Detection, Minimization and Analysis on the Set of Driver Nodes in Boolean Networks Control", Kyoto University, Japan | 3/2016 |
| "A Systematic Framework to Derive N-glycan Biosynthesis Process and the Automated Construction of Glycosylation Networks", Biotechnology and Bioinformatics Symposium (BIOT) 2015, Provo, USA | 12/2015 |
| "Novel Data Mining Techniques for Gene Expression Analysis, Genome Analysis and Drug Design", Kyoto University, Japan | 6/2015 |
| "On Mathematical Models and Data Mining Techniques for Bioinformatics", HKU | 1/2015 |
| "A Mathematical Model to Derive N-glycan Biosynthesis Networks", Soka University, Japan | 8/2014 |

RESEARCH EXPERIENCE

Professional Training

- | | |
|---|--------|
| The Machine Learning Summer School 2016 , University of Cadiz, Spain | 5/2016 |
| Intensively learnt courses such as kernel methods (by Arthur Gretton), stochastic optimization (Dr. Francis Bach), gaussian processes (Dr. Neil Lawrence), deep convolution neural networks (Dr. Stephane Mallet), deep learning (Dr. Oriol Vinyals), etc.. | |

Competitions

- | | |
|--|---------------|
| The American Medical Informatics Association Knowledge Discovery and Data Mining (AMIA KDDM) 2016 Data Competition | 6/2016-8/2016 |
| Predicted the insurgence of post-surgical infections in patients undergoing gastro-intestinal surgery based on feature encoding of blood test values over time, missing-rate-and-PCA-based feature selection, and cost sensitive Logistic Regression | |
| Mathematical Contest in Modelling (MCM) | 1/2012-2/2012 |
| Designed MATLAB-based SOM neural network model to describe and classify leaves of a tree with image pixels. (awarded as the <i>Honorable Mention</i>) | |
| Chinese Mathematical Contest in Modelling | 8/2011-9/2011 |
| Analyzed heavy metal pollution on urban topsoil through cubic interpolation & k-means clustering; located pollution sources by binary image erosion algorithm. (awarded as the <i>Second Award</i>) | |

Student Projects

- | | |
|---|---------------|
| Sun Yat-sen University Undergraduate Research Projects No. 20110118 | 4/2012-4/2013 |
| Applied double buffering mechanism into dynamic encryption system for the investigation on encrypt on-the-fly | |
| Guangdong Province Undergraduate Innovative Projects No. 1055811118 | 6/2011-4/2012 |

Established multi-target programming models for linear crossing, cross intersection, and T section for intelligent traffic light design- Haizhu District as a case study. (awarded as an *Excellent Project*)

TEACHING

Teaching Assistant, Department of Mathematics, The University of Hong Kong
(Give supplement lectures, exercise & programming lessons twice per week)
MATH 1011 University Mathematics (**1** semester)
MATH 2601&3601 Numerical Analysis (**2** semesters)
MATH 1211&2211 Multivariable Calculus (**3** semesters)

9/2013-2/2017

ACADEMIC SERVICES

Member

International Society for Computational Biology (ISCB), 2018-2019
The University of Hong Kong Shenzhen Institute of Research and Innovation (HKU-SIRI), 2016-2017
Society for Industrial and Applied Mathematics (SIAM), 2014-2018

Reviewer

Bioinformatics
Genome Biology
IEEE Transactions on Neural Networks and Learning Systems (TransNNLS)
Nature Communications
Neurocomputing
The First CCF Bioinformatics Conference (CBC 2016)
IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016)
The 15th Asia Pacific Bioinformatics Conference (APBC 2017)
The 10th International Conference on Systems Biology (ISB 2016)

Judge

2019 Women in STEM Symposium Poster Judge, Baltimore, USA, 04/06/2019
2018 Women in STEM Symposium Poster Judge, Baltimore, USA, 04/07/2018

CORE TECHNICAL SKILLS

Programming	R (9 years), MATLAB (8 years), C,C++, python, LINGO/LINDO, SPSS
Language	Mandarin & Cantonese & Hakka (native), English (working proficiency), Japanese (elementary)
Coursera	Machine Learning; Big Data, Genes, and Medicine; Finding Hidden Messages in DNA; Genome & Sequencing