

Wenpin Hou, Ph.D.

Date of Preparation: April 30, 2023

PERSONAL DATA

Name: Wenpin Hou
Contact Information: Room 650, 722 West 168th Street, New York, NY 10032, USA

ACADEMIC APPOINTMENTS / WORK EXPERIENCE

Assistant Professor (*tenure-track*), Columbia University, New York, NY, USA 7/2022-Present
Department of Biostatistics, Mailman School of Public Health

EDUCATION

Ph.D., Mathematics, The University of Hong Kong, Hong Kong, China 9/2013-8/2017
Mentor: Dr. Wai-Ki Ching
Thesis: Mathematical modelling and optimization in biological networks and data
Degree Awarded Date: November 25, 2017
B.Sc., Information and Computational Science, Sun Yat-sen University, Guangzhou, China 9/2009-7/2013
Advisor: Dr. Tianshou Zhou
Degree Awarded Date: June 19, 2013

TRAINING

Postdoctoral Fellow

Johns Hopkins University, Department of Biostatistics, Baltimore, MD, USA 8/2019-06/2022
Mentors: Drs. Stephanie C. Hicks and Hongkai Ji
K99 Phase Mentors: Drs. Hongkai Ji, Stephanie C. Hicks and Andrew P. Feinberg
K99 Phase Advisors: Drs. Xiaobin Wang, Kasper Hansen and Gregory Hager
Johns Hopkins University, Department of Computer Science, Baltimore, MD, USA 8/2017-8/2019
Mentors: Drs. Aravinda Chakravarti and Suchi Saria

Visiting Researcher

NYU Langone Health, Center for Human Genetics and Genomics, New York, NY, USA 12/2018-12/2018
MD Anderson Cancer Center, Department of Bioinformatics & Computational Biology, Houston, TX, USA 7/2017-8/2017
Kyoto University, Bioinformatics Center, Kyoto, Japan 5/2016-8/2016
Kyoto University, Bioinformatics Center, Kyoto, Japan 5/2015-8/2015
Soka University, Bioinformatics Lab, Tokyo, Japan 6/2014-8/2014

HONORS AND AWARDS

Emerging Leaders in Computational Oncology Award, Memorial Sloan Kettering Cancer Center 2021
NIH Pathway to Independence Award (K99/R00), NIH/NHGRI 2021
2021 Women in Statistics and Data Science Conference Award, American Statistical Association 2021
Honorary Research Associate in Mathematics, The University of Hong Kong 2018-2020
First-place Winning Poster at the 13th Annual Symposium and Poster Session on Genomics and Bioinformatics 2019
University Postgraduate Fellowships, Philip K H Wong Foundation, The University of Hong Kong 2013-2017
Postgraduate Scholarship, The University of Hong Kong 2013-2017
Excellent Teaching Award, Department of Mathematics, The University of Hong Kong 2016
Doris Chen Postgraduate Travel Grants 2016
Xu Chongqing Award for Academic Excellence (*two students awarded among Philosophy, Math, Physics, and Chemistry Schools*) 2012
Academic Award for Excellent Undergraduates (*annually awarded for three consecutive years*) 2010-2012

PROFESSIONAL ORGANIZATIONS AND SOCIETIES

Memberships

International Chinese Statistical Association (ICSA), 2023-2024
American Statistical Association (ASA), 2021-2022
Eastern North American Region (ENAR), 2019-2020
International Society for Computational Biology (ISCB), 2018-2019

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

Reviewer

Peer-reviewed Journals

Bioinformatics (4)
Communications Biology
Genomics, Proteomics & Bioinformatics
Genome Biology
IEEE Transactions on Neural Networks and Learning Systems
NAR Genomics and Bioinformatics
Nature Communications (3)
Nature Methods
Nature Portfolio
Neurocomputing
PLoS Computational Biology
Precision Nutrition

Conference

The 12th International Chinese Statistical Association Conference ([ICSA2023](#))
International Conference on Health Policy Statistics ([ICHPS 2023](#))
The 15th Asia Pacific Bioinformatics Conference (APBC 2017)
The First CCF Bioinformatics Conference (CBC 2016)
IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016)
The 10th International Conference on Systems Biology (ISB 2016)

Judge

Poster Judge, 2019 Women in STEM Symposium, Baltimore, USA, 4/6/2019
Poster Judge, 2018 Women in STEM Symposium, Baltimore, USA, 4/7/2018

FELLOWSHIP AND GRANT SUPPORT

Present Support

K99/R00HG011468, NIH/NHGRI

3/2021-3/2026

Title: Computational methods for inferring single-cell DNA methylation and its spatial landscape

Role: Principal Investigator

Direct Cost: \$ 617, 406

Pending Support

R35GM150887, NIH/NIGMS

7/2023-6/2028

Title: Methods for inferring and analyzing gene regulatory networks using single-cell multiomics and spatial genomics data

Role: Principal Investigator

Direct Cost: \$ 1, 249, 998

Status: Submitted on 09/30/2022, impact score 34, pending Advisory Council Review

1R01MH133561-01, Research Foundation for Mental Hygiene, Boldrini, M., sub to Mann, John

07/2023-06/2028

Title: Human brain multi-omics to decipher major depression pathophysiology

Role: Collaborator (PI: Mann, John)

Total Cost: \$ 560,173

Status: Submitted on 09/29/2022, impact score 33, pending Advisory Council Review

1R01AR082380-01A1, NIH/NIAMS

9/2023-8/2028

Title: Identifying extrinsic regulation of skin-specific vasculature in immune surveillance using biomimetic skin models

Role: Collaborator (PI: Abaci, Erbil)

Total Cost: \$ 2,666,024

Status: Submitted on 3/6/2023

104413, NSF

10/2023-9/2026

Title: Collaborative research: robust perturbation and objective-based experimental design in regulatory networks under uncertainty

Role: Principal Investigator (sub from Imani, Mahdi)

Total Cost: \$ 444,779

Status: Submitted on 3/22/2023

EDUCATIONAL CONTRIBUTIONS

Classroom Teaching with *Certificate of Teaching and Learning in Higher Education*

Teaching Assistant and Lab Instructor (received *Excellent Teaching Award*), Department of Mathematics, The University of Hong Kong, Hong Kong, China 9/2013-2/2017

Responsibility: give supplement lectures, programming lessons, and exercise class twice per week; design and grade assignments; provide instruction to individual or small groups of students to improve academic performance; advise students about their academic and career development

- MATH 1011 University Mathematics (1 semester)
- MATH 2601&3601 Numerical Analysis (2 semesters)
- MATH 1211&2211 Multivariable Calculus (3 semesters)

Mentorship

Master's Students, Department of Biostatistics, Columbia University, New York, NY, USA

- Jingyi Yao (Theory and Methods Track) 3/2023-Present
- Tianchuan Gao (Theory and Methods Track), upcoming PhD student at Indiana University Department of Biostatistics and Health Data Science 9/2022-Present
- Wenhan Bao (Theory and Methods Track), upcoming PhD student at Florida University Department of Biostatistics 8/2022-Present

Community Education

- Harvard Medical School - Chinese Students and Scholars Association (HMS-CSSA) Career Development Series, Virtual 8/19/2022
- Chan Zuckerberg Initiative (CZI) Faculty Search Bootcamp: Recent Jobseeker Panel, Virtual 8/8/2022

PUBLICATIONS *The * indicates the corresponding author(s).*

In Preprint:

1. Hou, W., and Ji, Z. 2023. Reference-free and cost-effective automated cell type annotation with GPT-4 in single-cell RNA-seq analysis. [bioRxiv, 2023 April 21](#). In Journal Review.
2. Jackson, C., Cherry, C., Bom, S., Dykema, A., Thompson, E., Zheng, M., Ji, Z., Hou, W., Li, R., Zhang, H. and Choi, J., Rodriguez, F., Weingart, J., Yegnasubramanian, S., Lim, M., Bettegowda, C., Powell, J., Eliesseff, J., Ji, H., and Pardoll, D. 2023. Distinct myeloid derived suppressor cell populations promote tumor aggression in glioblastoma. [bioRxiv, 2023 Jan 1](#). In Journal Review.
3. Hou, W. and Ji, Z., 2023. GeneTuring tests GPT models in genomics. [bioRxiv, pp.2023-03](#). In Journal Review.
4. Wang, W., Wang, Y., Liu, D., Hou, W., Zhou, T. and Ji, Z., 2022. GeneSegNet: A deep learning framework for cell segmentation by integrating gene expression and imaging. [bioRxiv, pp.2022-12](#). Software package: [GeneSegNet](#). In Journal Review.
5. Hou, W. and Ji, Z., 2022. Decomposing spatial heterogeneity of cell trajectories with Paella. [bioRxiv, pp.2022-09](#). Software package: [Paella](#). In Journal Review.
6. 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, 2021 Jul 12](#). PMID: 34282418. PMCID: PMC8288148. Software package: [Lamian](#). In Journal Revision.

Published:

Methods for Single-cell Genomics and Epigenomics

1. Hou, W. and Ji, Z., 2022. Palo: Spatially-aware color palette optimization for single-cell and spatial data. [Bioinformatics, June 01, 2022](#). Software package: [Palo](#). PMID: 35642896.
2. Hou, W. and Ji, Z., 2022. Single-cell unbiased visualization with SCUBI. [Cell Reports Methods, 100135, 2022](#). Software package: [SCUBI](#). PMID: 35224531. PMCID: PMC8871596.
3. Caushi, J.X., Zhang, J., Ji, Z., Vaghasia, A., Zhang, B., Hsiue, E., Mog, B., Hou, W., Justesen, S., Blosser, R., Tam, A., Anagnostou, V., Cottrell, T.R., Guo, H., Chan, H., Singh, D., Thapa, S., Dykema, A., Choudhury, C., Aparicio, L., Cheung, L., Lanis, M., Belcaid, Z., Asmar, M.E., Illei, P., Brock, M., Ha, J., Bush, E., Park, B., Bott, M., Naidoo, J., Marrone, K.A., Reuss, J.E., Velculescu, V.E., Chaff, J.E., Kinzler, K.W., Zhou, S., Vogelstein, B., Taube, J.M., Merghoub, T., Brahmer, J.R., Hellmann, M.D., Forde, P.M., Yegnasubramanian, S., Ji, H., Pardoll, D.M., Smith, K.N. (2021). Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. [Nature. 2021 Aug; 596\(7870\):126-132](#). PMID: 34290408. PMCID: PMC8338555.

Note 1: **Discussed in Nature Reviews Clinical Oncology** [Tumour antigen-induced T cell exhaustion – the archenemy of immune-hot malignancies](#)

Note 2: I contributed to the pseudotime analysis and functional analysis (Figure 4g-h, Extended data Figure 12b-f).

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. PMID: 32854757. PMCID: PMC7450705.
5. Ji, Z., Zhou, W., **Hou, W.** and Ji, H., 2020. SCATE: single-cell ATAC-seq signal extraction and enhancement. [Genome Biology](#), **21**, 161. PMID: 32620137. PMCID: PMC7333383. Software packages: [SCATE](#). [SCATEData](#).

Controllability in Boolean Networks

6. **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. PMID: 30543810.
7. **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. PMID: 30679639. PMCID: PMC6345816.

Glycosylation Networks

8. **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. PMID: 27454116. PMCID: PMC4965717.

Public Health for Low-income Minority

9. Huang, W., Igusa, T., Wang, G., Buckley, J.P., Hong, X., Bind, E., Steffens, A., Mukherjee, J., Haltmeier, D., Ji, Y., Xu, R., **Hou, W.**, Fan, Z., and Wang, X., 2022. In-utero co-exposure to toxic metals and micronutrients on childhood risk of overweight or obesity: new insight on micronutrients counteracting toxic metals. [International Journal of Obesity](#), **46**, 1435–1445.
10. **Hou, W.***, Zhang, M., Ji, Y., Hong, X., Wang, G., Xu, R., Liang, L., Saria, S. and Ji, H., 2022. A prospective birth cohort study of maternal prenatal cigarette smoking assessed by self-report and biomarkers on childhood risk of overweight or obesity. [Precision Nutrition](#), **1**(3), e00017.
11. 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.
12. 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 for Biomedical Data

13. 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. PMID: 29994681.
14. 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. PMID: 29322919. PMCID: PMC5763304.
15. 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.
16. 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 \(ISORA\)](#), pp.144-150, Luoyang, China, 2015.
17. 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 \(ISORA\)](#), pp.151-157, Luoyang, China, 2015.
18. 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 \(ISB\)](#), pp.88-93, Qingdao, China, 2014.

Thesis

19. **Hou, W.** (2017) Mathematical modelling and optimization in biological networks and data. Ph.D. Thesis. The University of Hong Kong.

SOFTWARE

1. **Paella**, R package, a software for decomposing spatial heterogeneity of cell trajectories
Github: <https://github.com/Winnie09/Paella>
Role: lead author, developer
2. **Lamian**, R package, a systematic framework for differential pseudotime analysis in multiple single-cell RNA-seq
Github: <https://github.com/Winnie09/Lamian>
Role: lead author, developer
3. **SCUBI**, R package, a software for single-cell unbiased visualization
Github: <https://github.com/Winnie09/scubi>
Role: lead author, co-developer
4. **SCATE**, R package, a software for single-cell ATAC-seq signal extraction and enhancement
Github: <https://github.com/Winnie09/SCATE>
Role: co-author, co-developer
5. **SCATEData**, R package, a software providing extensive data resources for the package SCATE.
Github: <https://github.com/Winnie09/SCATEData>
Role: co-author, co-developer
6. **Palo**, R package, a software for spatially-aware color palette optimization in single-cell and spatial data
Github: <https://github.com/Winnie09/Palo>
Role: lead author, co-developer

PRESENTATIONS

Seminar Talks

1. Johns Hopkins University, Department of Biostatistics, Statistical Genetics Working Group. Baltimore, MD, USA. 12/2021.
2. Johns Hopkins University, Department of Biostatistics. Baltimore, MD, USA. 3/2019.
3. MD Anderson Cancer Center, Department of Bioinformatics and Computational Biology. Houston, TX, USA. 7/2017
4. Kyoto University, Bioinformatics Center. Kyoto, Japan. 3/2016.
5. Kyoto University, Bioinformatics Center. Kyoto, Japan. 6/2015.
6. Soka University, Department of Bioinformatics. Tokyo, Japan. 8/2014.

Job Talks

1. University of Michigan, Department of Computational Medicine and Bioinformatics. 3/2022
2. Duke University, Department of Biostatistics and Bioinformatics. 2/2022
3. Columbia University, Department of Biostatistics. Virtual. 2/2022
4. Mount Sinai, Department of Genetics and Genomic Sciences. Virtual. 2/2022.
5. New York University, Department of Biostatistics. Virtual. 2/2022
6. UNC Chapel Hill, Department of Mathematics. Chapel Hill, NC, USA. 2/2022.
7. Weill Cornell Medicine, Department of Population Health Sciences. Virtual. 2/2022.
8. New York University, NYU Langone Institute for Systems Genetics. Virtual. 1/2022.
9. University of Pittsburgh, Department of Biological Sciences. Virtual. 12/2021.
10. University of Southern California, Division of Biostatistics at Keck School of Medicine. Virtual. 10/2021.
11. University of Southern California, Center for Craniofacial Molecular Biology. Virtual. 9/2021.

Conference Talks

1. Computational methods for analyzing gene expression and regulation dynamics at single-cell and spatial resolution. *Emerging Leaders Symposium in Computational Oncology*, Memorial Sloan Kettering Cancer Center, Virtual, 10/14-15/2021.
2. Lamian: a statistical framework for differential pseudotime analysis in multiple single-cell RNA-seq samples. *2021 Joint Statistical Meetings (JSM)*, Virtual, 8/8/2021.
3. A systematic evaluation of single-cell RNA-seq imputation methods. *ENAR 2020 Spring Meeting*, Virtual, 3/24/2020.
4. A systematic framework to derive N-glycan biosynthesis process and the automated construction of glycosylation networks. *Biotechnology and Bioinformatics Symposium (BIOT) 2015*, Provo, UT, USA, 12/10/2015.

Posters Sessions

1. A computational framework for differential pseudotime analysis across conditions with multiple single-cell RNA-seq samples reveals T cell immune dynamics associated with COVID-19 disease severity, *CSHL Systems Immunology 2021*, Virtual, 4/20-23/2021.
2. A systematic evaluation of single-cell RNA-seq imputation methods, *13th Annual Symposium and Poster Session on Genomics and Bioinformatics*, Johns Hopkins University, Baltimore, MD, USA, 10/17/2019. (awarded as *first-place winning poster*)
3. 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, NY, USA, 12/8-10/2018.
4. On orthogonal feature extraction model with applications in cancer prediction, [Machine Learning Summer School](#), University of Cadiz, Cadiz, Spain, 5/18/2016.