Wenpin Hou, Ph.D.

Date of Preparation: April 30, 2024

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
Affiliated Member, Data Science Institute, Columbia University, New York, NY, USA   1/2024-Present
Center for Foundations of Data Science
Center for Health Analytics

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: 11/25/2017
B.Sc., Information and Computational Science, Sun Yat-sen University, Guangzhou, China   9/2009-6/2013
Advisor: Dr. Tianshou Zhou
Degree Awarded Date: 6/19/2013

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. Suchi Saria and Aravinda Chakravarti

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

Maximizing Investigators Research Award (MIRA) for Early Stage Investigators, NIH/NIGMS	2023
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 (*received annually for three years, including one First-Class distinction*)	2010-2012

ADMINISTRATIVE LEADERSHIP AND ACADEMIC SERVICE

Academic Service
Co-found the Statistical Genetics and Genomics (SGG) Working Group   9/2022
- Form a group of 35 sign-up members and in charge of the mailing list
- Organize bi-weekly SGG seminars and invite six external speakers from UPenn, Duke, UMich, UNC Chapel Hill, NCI, Boston University, etc. from 9/2023 to 12/2024

Departmental Committees
- Member, PhD Admission Committee   10/2023-present
- Member, Application Qualifying Exams Committee   10/2023-present
- Member, Curriculum Committee   10/2023-present

Conference Award Committees
- The 12th International Chinese Statistical Association Conference (ICSA2023)
- International Conference on Health Policy Statistics (ICHPS 2023)

Administrative Leadership
- Seminar speaker, PhD student recruitment visit day   2/2024
- Panelist, Doctoral Job Search: Academia workshop, Columbia Mailman School of Public Health, Virtual   10/2023
- Oral presenter and poster judge, Columbia Biostatistics Annual Research Symposium (CBARS)   9/2023

PROFESSIONAL ORGANIZATIONS AND SOCIETIES

Memberships

Institute of Mathematical Statistics (IMS)	2023-2024
International Chinese Statistical Association (ICSA)	2023-2024
American Statistical Association (ASA)	2021-2022, 2023-2024
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

Journal Reviewer
Bioinformatics (4), Communications Biology, Genomics, Proteomics & Bioinformatics, Genome Biology (3), IEEE Transactions on Neural Networks and Learning Systems, NAR Genomics and Bioinformatics, Nature Communications (10), Nature Methods (3), Nature Portfolio, Neurocomputing, PLoS Computational Biology (4), Precision Nutrition

GRANT SUPPORT   *The* * *indicates the grants cannot be accepted currently because of over-funded status*

Present Support
R35GM150887 (R01-equivalent), NIH/NIGMS   9/2023-8/2028
Title: Methods for inferring and analyzing gene regulatory networks using single-cell multiomics and spatial genomics data
Role: Principal Investigator, 5.292 CM
Direct Cost: $1,249,998

K99/R00HG011468, NIH/NHGRI   3/2021-6/2025
Title: Computational methods for inferring single-cell DNA methylation and its spatial landscape
Role: Principal Investigator, 5.508 CM
Direct Cost: $617,406

1R01MH133561-01*, NIH/NIMH   9/2023-6/2028
Title: Human brain multi-omics to decipher major depression pathophysiology
Role: Co-Investigator, 0.6 CM (PI: Boldrini, M.)
Total Cost: $3,863,494

R01AG076949 Supplement*, NIH   12/2023-11/2024
Title: Harmonization and joint analysis of human brain single-cell datasets from neurotypical aging controls and alzheimer's disease patients
Role: Co-Investigator, 3 CM (PI: Boldrini, M.)
Total Cost: $250,000

EDUCATIONAL CONTRIBUTIONS

Direct Teaching

Instructor, Department of Biostatistics, Columbia University
- P8131 Biostatistical Methods II   1/2024-5/2025
  - Sole instructor for this mandatory course, providing three hours of instruction per week
  - Curriculum includes Generalized Linear Models, Longitudinal Data Analysis, and Survival Analysis
  - Taught 113 master's level students

Guest Lecturer, Department of Biostatistics, Columbia University
- Graduate Student Research Seminar (GSRS), Columbia Biostatistics Department   10/2/2023

Teaching Assistant and Lab Instructor, Department of Mathematics, The University of Hong Kong   9/2013-2/2017
Certificate: Certificate of Teaching and Learning in Higher Education
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)

Advising and Mentorship (The ‡ indicates that I am the primary mentor or practicum advisor.)
Postdoctoral Research Scientists, Department of Biostatistics, Columbia University, New York, NY, USA
- Won Eui Hong‡ (co-mentored by Dr. Ying Wei)   2/2024-Present

PhD Students, Department of Biostatistics, Columbia University, New York, NY, USA
- Elly Kipkogei (Provost diversity fellow), mentor through SPRIS II course   9/2023-1/2024

Master's Students, Department of Biostatistics, Columbia University, New York, NY, USA
- Yifei Zhao‡, (Y2022, Theory and Methods Track)   10/2023-Present
- Lehan Zou (Y2023,Theory and Methods Track)   9/2023-Present
- Aiying Huang (Y2023, Public Health Data Science track)   9/2023-Present
- Jingyi Yao‡ (Y2022,Theory and Methods Track)   3/2023-Present
- Yiying Wu (Y2023,Theory and Methods Track)   7/2023-8/2023
- Tianchuan Gao‡ (Y2021, Theory and Methods Track), current PhD student at Indiana University   9/2022-5/2023
- Wenhan Bao (Y2021, Theory and Methods Track), current PhD student at Florida University   8/2022-6/2023

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

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

Methods for Single-cell Genomics and Epigenomics

1. **Hou, W.***, and Ji, Z.*, 2024. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. Nature Methods, 2024 March 25. **[IF: 48]**. Software package: GPTCelltype.
   Note 1: Featured in 12 news outlets, such as Columbia News Spotlight, Columbia MSPH News, Science Daily, The Medical News, Health Tech World, etc.
   Note 2: This research output has an Altmetric Attention Score of 284. It ranked the top 1 when comparing to 75 others from the same source and published within six weeks.
2. **Hou, W.**, Ji, Z., Chen, Z., Wherry, E.J., Hicks, S.C. and Ji, H., 2023. A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. Nature Communications 14, 7286. Software package: Lamian.
3. Dykema, A.G., Zhang, J., Cheung, L.S., Connor, S., Zhang, B., Zeng, Z., Cherry, C.M., Li, T., Caushi, J.X., Nishimoto, M., Munoz, A.J., Ji, Z., **Hou, W.**, Zhan, W., Singh, D., Zhang, T., Rashid, R., Mitchell-Flack, M., Bom, S., Tam, A., Ionta, N., Aye, T.H.K., Wang, Y., Sawosik, C.A., Tirado, L.E., Tomasovic, L.M., Spangler, J.B., Anagnostou, W., Yang, S., Spicer, J., Rayes, R., Taube, J., Brahmer, J.R., Forde, P.M., Yegnasubramanian, S.*, Ji, H.*, Pardoll, M.*, and Smith K.N.*, 2023. Lung tumor–infiltrating Treg have divergent transcriptional profiles and function linked to checkpoint blockade response. Science Immunology, 8(87). PMID: 37713507. PMCID: PMC10629528.

4. Wang, Y., Wang, W., Liu, D., **Hou, W.**, Zhou, T. and Ji, Z., 2023. GeneSegNet: A deep learning framework for cell segmentation by integrating gene expression and imaging. Genome Biology 24, 235. Software package: GeneSegNet.

5. **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.

6. **Hou, W.** and Ji, Z., 2022. Single-cell unbiased visualization with SCUBI. Cell Reports Methods, 100135, 2022. Software package: SCUBI. PMID: 35224531. PMCID: PMC8871596.

7. 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., Chaft, 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 596, 126–132. PMCID: PMC8338555.
Note 1: Discussed in Tumour antigen-induced T cell exhaustion – the archenemy of immune-hot malignancies in Nature Reviews Clinical Oncology 18, 749–750 (2021).
Note 2: I contributed to the pseudotime analysis and functional analysis (Figure 4g-h, Extended data Figure 12b-f).

8. **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.

9. 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

10. **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.

11. **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

12. **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

13. Xu, R., Hong, X., Ladd-Acosta, C., Buckley, J.P., Choi, G., Wang, G., **Hou, W.**, Wang, X., Liang, L. and Ji, H., 2023. Contrasting Association of Maternal Plasma Biomarkers of Smoking and 1-Carbon Micronutrients with Offspring DNA Methylation: Evidence of Aryl Hydrocarbon Receptor Repressor Gene–Smoking–Folate Interaction. The Journal of Nutrition, 153(8), 2339-2351.

14. 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.

15. **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 ciga-rette smoking assessed by self-report and biomarkers on childhood risk of overweight or obesity. Precision Nutrition, 1(3), e00017.

16. 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.

17. 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

18. 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.

19. 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.
20. 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.
21. 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.
22. 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.
23. 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

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

Posters

25. CSHL Systems Immunology 2021, Virtual. "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." 4/2021
26. 13th Annual Symposium and Poster Session on Genomics and Bioinformatics, Johns Hopkins University, Baltimore, MD, USA. "A systematic evaluation of single-cell RNA-seq imputation methods." 10/2019 (awarded as *first-place winning poster*)
27. 11th annual RECOMB/ISCB Conference on Regulatory & Systems Genomics with DREAM Challenges, New York University, New York, NY, USA. "Causal gene regulatory network construction using single-cell RNA-seq and single-cell ATAC-seq data." 12/2018
28. Machine Learning Summer School, University of Cadiz, Cadiz, Spain. "On orthogonal feature extraction model with applications in cancer prediction." 5/2016

PRESENTATIONS

Invited Seminar Talks

1. Memorial Sloan-Kettering Cancer Center, Department of Epidemiology and Biostatistics. New York, USA. 6/2024 (scheduled)
2. Statistical Genetics and Genomics Working Group at Columbia Biostatistics. "Statistical Methods for Deciphering Cellular Trajectories in Single-cell and Spatial Transcriptomics Data." 9/2023
3. Johns Hopkins University, Department of Biostatistics, Statistical Genetics Working Group. Baltimore, MD, USA. 12/2021
4. Johns Hopkins University, Department of Biostatistics. Baltimore, MD, USA. 3/2019
5. MD Anderson Cancer Center, Department of Bioinformatics and Computational Biology. Houston, TX, USA. 7/2017
6. Kyoto University, Bioinformatics Center. Kyoto, Japan. 3/2016
7. Kyoto University, Bioinformatics Center. Kyoto, Japan. 6/2015
8. Soka University, Department of Bioinformatics. Tokyo, Japan. 8/2014

Invited Panel Discussion

1. Adapting to AI in Healthcare: an in-Person Workshop, Columbia University Irving Medical Center. "Open AI in Basic & Translational Ressearch — Opportunities in Health & Diseases". 4/2024

Invited 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. 2/2022
4. Mount Sinai, Department of Genetics and Genomic Sciences. 2/2022
5. New York University, Department of Biostatistics. 2/2022
6. UNC Chapel Hill, Department of Mathematics. Chapel Hill, NC, USA. 2/2022
7. Weill Cornell Medicine, Department of Population Health Sciences. 2/2022

8. New York University, NYU Langone Institute for Systems Genetics. 1/2022
9. University of Pittsburgh, Department of Biological Sciences. 12/2021
10. University of Southern California, Division of Biostatistics at Keck School of Medicine. 10/2021
11. University of Southern California, Center for Craniofacial Molecular Biology. 9/2021

Conference Talks

Invited:
1. Annual Synthetic Regeneration Symposium, Columbia University. New York, USA. "Differential pseudotime analysis and spatio-temporal analysis in single-cell data". 7/2023
2. 2023 ICSA Applied Statistics Symposium. University of Michigan Ann Arbor, USA. "Statistical Methods Development in Single-cell Genomics and Multi-omics Data. " 6/2023
3. Columbia Stem Cell Initiative (CSCI) Retreat. Edith Macy Retreat Center, Briarcliff Manor, USA. "Computational methods for spatial and temporal trajectory analysis in single-cell data." 5/2023

Contributed:
1. ENAR 2024 Spring Meeting, Baltimore, USA. "Decomposing spatial heterogeneity of cell trajectories with Paella." 3/2024
2. Emerging Leaders Symposium in Computational Oncology, Memorial Sloan Kettering Cancer Center, Virtual. "Computational methods for analyzing gene expression and regulation dynamics at single-cell and spatial resolution." 10/2021
3. 2021 Joint Statistical Meetings (JSM), Virtual. "Lamian: a statistical framework for differential pseudotime analysis in multiple single-cell RNA-seq samples." 8/2021
4. ENAR 2020 Spring Meeting, Virtual. "A systematic evaluation of single-cell RNA-seq imputation methods." 3/2020
5. Biotechnology and Bioinformatics Symposium (BIOT) 2015, Provo, UT, USA. "A systematic framework to derive *N*-glycan biosynthesis process and the automated construction of glycosylation networks." 12/2015