

Siyuan Ma, Ph.D.

Assistant Professor, Department of Biostatistics, Vanderbilt University Medical Center
2525 West End Avenue · Room 1140 · Nashville, TN 37203
✉ siyuan.ma@vumc.org 🌐 [syma-research.github.io](https://github.com/syma-research) | Updated: 24 Oct 2024

Academic Appointment

Vanderbilt University Medical Center
Assistant Professor of Biostatistics 2022-

Education

Harvard T.H. Chan School of Public Health, Ph.D. in Biostatistics 2019
Thesis: Statistical Methods for Population Structure Discovery in Meta-Analyzed 'Omics Studies
Advisor: Curtis Huttenhower, Ph.D.
Harvard T.H. Chan School of Public Health, M.S. in Biostatistics 2015
Peking University, B.S. in Statistics 2013

Professional Experience

University of Pennsylvania, Postdoctoral Researcher 2020-2022
Advisor: Hongzhe Li, Ph.D.
Harvard T.H. Chan School of Public Health, Postdoctoral Researcher 2019-2020
Advisors: Curtis Huttenhower, Ph.D., Lucas Janson, Ph.D.

Research Grants as Principal Investigator

NIH Environmental influences on Child Health Outcomes (ECHO) Program Opportunity Opportunities and Innovation Fund (PI: Ma) 2024-2026
The Role of Airway "Sociobiome" in Neighborhood-Related Childhood Asthma Burden
\$200,000 (Total Cost)
VICC SPORE in Gastrointestinal Cancer Career Enhancement Program (PI: Ma) 2023-2024
Computational Profiling of Protein Colocalization for Immunotherapy Targets in Colorectal Cancer, with Spatial Omics
\$50,000 (Direct Cost)

Publications

First-/senior-authorships are **bolded**. * indicates shared authorship.

1. Yan, B., Nam, Y., Li, L., Deek, R.A., Li, H.* and **Ma, S.***, 2024. Recent advances in deep learning and language models for studying the microbiome. *arXiv preprint*, arXiv:2409.10579.
2. Gao, Z.*, Li, L.*, **Ma, S.***, Wang, Q., Hemphill, L. and Xu, R., 2024. Examining the potential of ChatGPT on biomedical information retrieval: fact-checking drug-disease associations. *Annals of Biomedical Engineering*, 52(8), pp.1919-1927.
3. Mei, Z., Wang, F., Bhosle, A., Dong, D., Mehta, R., Ghazi, A., Zhang, Y., Liu, Y., Rinott, E., Ma, S. and Rimm, E.B., 2024. Strain-specific gut microbial signatures in type 2 diabetes identified in a cross-cohort analysis of 8,117 metagenomes. *Nature Medicine*, 30(8), pp.2265-2276.
4. Deek, R.A., Ma, S., Lewis, J. and Li, H., 2024. Statistical and computational methods for integrating microbiome, host genomics, and metabolomics data. *eLife*, 13, p.e88956.
5. Xiong, J., Kaur, H., Heiser, C.N., McKinley, E.T., Roland, J.T., Coffey, R.J., Shrubsole, M.J., Wrobel, J., Ma, S., Lau, K.S. and Vandekar, S., 2024. GammaGateR: semi-automated marker gating for single-cell multiplexed imaging. *Bioinformatics*, p.btae356.
6. Li, L., Zhou, J., Gao, Z., Hua, W., Fan, L., Yu, H., Hagen, L., Zhang, Y., Assimes, T.L., Hemphill, L. and Ma, S., 2024. A scoping review of using Large Language Models (LLMs) to investigate Electronic Health Records (EHRs). *arXiv preprint*, arXiv:2405.03066.
7. **Ma, S.**, Huttenhower, C. and Janson, L., 2024. Compositional Differential Abundance Testing: Defining and Finding a New Type of Health-Microbiome Associations. *bioRxiv preprint*, pp.2024-06.
8. Chatterjee, S., Fadikar, A., Hanumesh, V., Meshram, S.S., Zoh, R.S., Ma, S., Arunkumar, G. and Mallick, H., 2024. Group Heteroscedasticity-A Silent Saboteur of Power and False Discovery in RNA-Seq Differential Expression. *bioRxiv*, pp.2024-04.
9. Thompson, K.N., Bonham, K.S., Ilott, N.E., Britton, G.J., Colmenero, P., Bullers, S.J., McIver, L.J., Ma, S., Nguyen, L.H., Filer, A. and Brough, I., 2023. Alterations in the gut microbiome implicate key taxa and metabolic pathways across inflammatory arthritis phenotypes. *Science translational medicine*, 15(706), p.eabn4722.
10. **Ma, S.** and Li, H., 2023. A tensor decomposition model for longitudinal microbiome studies. *The Annals of Applied Statistics*, 17(2), pp.1105-1126.
11. Tan, Y., Shilts, M.H., Rosas-Salazar, C., Puri, V., Fedorova, N., Halpin, R.A., Ma, S., Anderson, L.J., Peebles Jr, R.S., Hartert, T.V. and Das, S.R., 2023. Influence of Sex on Respiratory Syncytial Virus Genotype Infection Frequency and Nasopharyngeal Microbiome. *Journal of Virology*, 97(3), pp.e01472-22.
12. **Ma, S.** and Li, H., 2022. Statistical and Computational Methods for Microbial Strain Analysis. In *Statistical Genomics* (pp. 231-245). New York, NY: Springer US.
13. Thompson, K.N., Oulhote, Y., Weihe, P., Wilkinson, J.E., Ma, S., Zhong, H., Li, J., Kristiansen, K., Huttenhower, C. and Grandjean, P., 2022. Effects of Lifetime Exposures to

Environmental Contaminants on the Adult Gut Microbiome. *Environmental Science & Technology*, 56(23), pp.16985-16995.

14. **Ma, S.**, Shungin, D., Mallick, H., Schirmer, M., Nguyen, L.H., Kolde, R., Franzosa, E., Vlamakis, H., Xavier, R.* and Huttenhower, C.*, 2022. Population structure discovery in meta-analyzed microbial communities and inflammatory bowel disease using MMUPHin. *Genome biology*, 23(1), pp.1-31.
15. Kondo, A., Ma, S., Lee, M.Y., Ortiz, V., Traum, D., Schug, J., Wilkins, B., Terry, N.A., Lee, H. and Kaestner, K.H., 2021. Highly multiplexed image analysis of intestinal tissue sections in patients with inflammatory bowel disease. *Gastroenterology*, 161(6), pp.1940-1952.
16. Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren, B., Schwager, E.H. and Chatterjee, S., 2021. Multivariable association discovery in population-scale meta-omics studies. *PLoS computational biology*, 17(11), p.e1009442.
17. **Ma, S.**, Ren, B., Mallick, H., Moon, Y.S., Schwager, E., Maharjan, S., Tickle, T.L., Lu, Y., Carmody, R.N., Franzosa, E.A. and Janson, L., 2021. A statistical model for describing and simulating microbial community profiles. *PLoS computational biology*, 17(9), p.e1008913.
18. Dantzer, K.W., Ma, S.*, Ngotho, P.*, Stone, W.J., Tao, D., Rijpmma, S., De Niz, M., Nilsson Bark, S.K., Jore, M.M., Raaijmakers, T.K. and Early, A.M., 2019. Naturally acquired immunity against immature Plasmodium falciparum gametocytes. *Science translational medicine*, 11(495), p.eaav3963.
19. **Ma, S.***, Ogino, S.*, Parsana, P., Nishihara, R., Qian, Z., Shen, J., Mima, K., Masugi, Y., Cao, Y., Nowak, J.A. and Shima, K., 2018. Continuity of transcriptomes among colorectal cancer subtypes based on meta-analysis. *Genome biology*, 19(1), pp.1-14.
20. Obaldia III, N., Meibalan, E., Sa, J.M., Ma, S., Clark, M.A., Mejia, P., Moraes Barros, R.R., Otero, W., Ferreira, M.U., Mitchell, J.R. and Milner, D.A., 2018. Bone marrow is a major parasite reservoir in Plasmodium vivax infection. *MBio*, 9(3), pp.e00625-18.
21. De Niz, M., Meibalan, E., Mejia, P., Ma, S., Brancucci, N.M., Agop-Nersesian, C., Mandt, R., Ngotho, P., Hughes, K.R., Waters, A.P. and Huttenhower, C., 2018. Plasmodium gametocytes display homing and vascular transmigration in the host bone marrow. *Science advances*, 4(5), p.eaat3775.
22. Mallick, H.*, **Ma, S.***, Franzosa, E.A., Vatanen, T., Morgan, X.C. and Huttenhower, C., 2017. Experimental design and quantitative analysis of microbial community multiomics. *Genome biology*, 18(1), pp.1-16.
23. Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A.A., Ren, B., Amir, A., Schwager, E., Crabtree, J., Ma, S., Abnet, C.C. and Knight, R., 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nature biotechnology*, 35(11), pp.1077-1086.
24. Pelle, K.G., Oh, K., Buchholz, K., Narasimhan, V., Joice, R., Milner, D.A., Brancucci, N.M., Ma, S., Voss, T.S., Ketman, K. and Seydel, K.B., 2015. Transcriptional profiling defines dynamics of parasite tissue sequestration during malaria infection. *Genome medicine*, 7, pp.1-20.

Invited Talks and Presentations

1. Testing for Immune Colocalization in Multiplexed Immunofluorescence Data. August 2024. Joint Statistical Meetings. Portland, OR.
2. Testing for Immune Colocalization in Multiplexed Immunofluorescence Data. July 2024. International Conference on Frontiers of Data Science. Hangzhou, China.
3. CompDA for Novel Health-Microbiome Associations and Controlling False Discoveries. June 2024. ICSA China Conference. Wuhan, China.
4. CompDA for Novel Health-Microbiome Associations and Controlling False Discoveries. June 2024. ICSA Applied Statistics Symposium. Nashville, TN.
5. CompDA for Novel Health-Microbiome Associations and Controlling False Discoveries. May 2024. STATGEN Conference. Pittsburgh, PA.
6. CompDA for Novel Health-Microbiome Associations and Controlling False Discoveries. March 2024. MCBIOS Conference. Atlanta, GA.
7. Flexible GLMM for Testing Cell Colocalization in Spatial Immunofluorescence Data. January 2024. Pacific Symposium on Biocomputing. Kailua-Kona, HI.
8. CompDA for Novel Health-Microbiome Associations and Controlling False Discoveries. December 2023. CMStatistics Conference. Virtual Seminar.
9. Modeling the Joint Distribution of Compositional Microbiome Data. June 2023. NESS Conference. Boston, MA.
10. Testing for Immune Colocalization in Multiplexed Immunofluorescence Data. May 2023. Statistical Methods in Imaging Conference, Minneapolis, MN.
11. Modeling the Joint Distribution of Compositional Microbiome Data. May 2023. Sydney Precision Bioinformatics, The University of Sydney, Virtual Seminar.
12. Modeling the Joint Distribution of Compositional Microbiome Data. March 2023. Department of Biostatistics, McGill University, Virtual Seminar.
13. Modeling the Joint Distribution of Compositional Microbiome Data. February 2023. Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Virtual Seminar.
14. A Tensor Factorization Model for Longitudinal Microbiome Data. December 2022. CM-Statistics Conference. Virtual Seminar.
15. Modeling the Joint Distribution of Compositional Microbiome Data. December 2022. Biostatistics Branch, National Cancer Institute. Virtual Seminar.
16. Modeling the Joint Distribution of Compositional Microbiome Data. August 2022. Department of Statistics, University of Pittsburgh. Virtual Seminar.
17. A Tensor Factorization Model for Longitudinal Microbiome Data. August 2022. Joint Statistical Meetings. Washington, DC.
18. A Tensor Factorization Model for Longitudinal Microbiome Data. May 2022. New England Statistics Symposium. Virtual Seminar.

19. Modeling the Joint Distribution of Compositional Microbiome Data. February 2022. Department of Biostatistics, Columbia University. Virtual Seminar.
20. Modeling the Joint Distribution of Compositional Microbiome Data. February 2022. Department of Biostatistics, New York University School of Global Public Health. Virtual Seminar.
21. Modeling the Joint Distribution of Compositional Microbiome Data. January 2022. Department of Biostatistics, Vanderbilt University Medical Center. Virtual Seminar.
22. A Statistical Model for Simulating and Testing for Microbiomes. October 2021. Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania. Virtual Seminar.
23. Testing Cell-Cell Interactions in Imaging Mass Cytometry Data. February 2021. Perelman School of Medicine, University of Pennsylvania. Virtual Seminar.
24. Meta-Analysis of Population Structure in the IBD Gut Microbiome. February 2019. Microbiome Working Group, Harvard T.H. Chan School of Public Health. Boston, MA.
25. Meta-Analysis of Population Heterogeneity in IBD Patients' Gut Microbiome. October 2017. The Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology. Cambridge, MA.
26. Subtype Continuity Revealed by Meta-analysis of the Colorectal Cancer Transcriptome. October 2015. Department of Biostatistics, Dana-Farber Cancer Institute. Boston, MA.

Softwares

microTensor: Tensor Decomposition of Longitudinal Microbiome Studies. github.com/syma-research/microTensor

SparseDOSSA 2: A Statistical Model for Describing and Simulating Microbial Community Profiles. github.com/biobakery/SparseDOSSA2

MMUPHin: Meta-analysis Methods with Uniform Pipeline for Heterogeneity in Microbiome Studies. www.bioconductor.org/packages/release/bioc/html/MMUPHin.html

Mentorship

Haoyue Li (PhD student), 2024-

Seyde Martinez (undergraduate student, summer intern), 2024

Eric Koplin (postdoctoral researcher), 2024-

Alexander Lin (undergraduate student), 2023-

Chih-Ting Yang (PhD student), 2023-

Tracy Brown (undergraduate student, summer intern), 2023

Teaching

Instructor. BIOS 6341-01: Fundamentals of Probability. Vanderbilt University Medical Center.

Guest Lecturer. BST 272: Introduction to Programming. Harvard T.H. Chan School of Public Health.

Teaching assistant. BST 235: Advanced Regression. Harvard T.H. Chan School of Public Health.

Teaching assistant. BST 254: Special Topics in Biostatistics (Structural Equations and Causal Inference). Harvard T.H. Chan School of Public Health.

Teaching assistant. BST 244: Analysis of Failure Time Data. Harvard T.H. Chan School of Public Health.

Teaching assistant. BST 215: Linear and longitudinal regression. Harvard T.H. Chan School of Public Health.

Teaching assistant. BST 508: Genomic data manipulation. Harvard T.H. Chan School of Public Health.

Professional Activities

SERVICE AND ENGAGEMENT

American Statistical Association, Section in Statistical Genetics and Genomics

Co-chair, Membership Engagement Committee, 2024-

Member, Membership Engagement Committee, 2023-2024

International Chinese Statistical Association

Co-chair, Student Paper Award Committee, 2024 ICSA Applied Statistics Symposium

MEMBERSHIP

American Statistical Association 2017-

Eastern North America Region of the IBS 2019-

International Society for Computational Biology 2023-

International Chinese Statistical Association 2024-

JOURNAL REFEREE

Nature communications, mSystems, Biometrics, The annals of applied statistics, PLOS Computational Biology, Bioinformatics, Frontiers in genetics, GigaScience, Journal of molecular biology, Plos One