

# Tiffany Tang

✉ ttang4@nd.edu • 🌐 tiffanymtang.github.io/

## ACADEMIC APPOINTMENTS

- University of Notre Dame**, Notre Dame, IN July 2024 – Present  
 Clare Boothe Luce Assistant Professor  
 Department of Applied and Computational Mathematics and Statistics (ACMS)
- University of Michigan**, Ann Arbor, MI July 2023 – July 2024  
 Postdoctoral Research Fellow (advised by Ji Zhu and Liza Levina)

## EDUCATION

- University of California, Berkeley**, Berkeley, CA July 2023  
 Ph.D. in Statistics (advised by Bin Yu)  
*Evelyn Fix Prize*
- Rice University**, Houston, TX May 2018  
 Bachelor of Science in Mathematics, Bachelor of Arts in Statistics  
*Summa cum laude, Distinction in Research and Creative Work*

## PAPERS

### Peer-Reviewed

- Duncan, J.\*, **Tang, T. M.\***, Elliott, C. F., Boileau, P., and Yu, B. [simChef: High-quality data science simulations in R](#). *Journal of Open Source Software*. 2024.
- Irajizad, E., Kenney, A. M.<sup>†</sup>, **Tang, T. M.<sup>†</sup>**, Vykoukal, J., Wu, R., Murage, E., Dennison, J. B., Sans, M., Long, J. P., Loftus, M., Chabot, J. A., Kluger, M. D., Kastrinos, F., Brais, L., Babic, A., Jajoo, K., Lee, L. S., Clancy, T. E., Ng, K., Bullock, A., Genkinger, J. M., Maitra, A., Do, K., Yu, B., Wolpin, B. M., Hanash, S., and Fahrman, J. F. [A blood-based metabolomic signature predictive of risk for pancreatic cancer](#). *Cell Reports Medicine*. 2023.
- Lu, Y., Fridlyand, J., **Tang, T. M.**, Qi, T., Simon, N., and Leng, N. [The Future will be Different than Today: Model Evaluation Considerations when Developing Translational Clinical Biomarker](#). *KDD Health Day - DSHealth Workshop*. 2021.
- Singh, C., Nasser, K., Tan, Y., **Tang, T. M.**, and Yu, B. [imodels: a python package for fitting interpretable models](#). *Journal of Open Source Software*. 2021.
- Tang, T. M.** and Allen, G. I. [Integrated Principal Components Analysis](#). *Journal of Machine Learning Research*. 2021.
- Baker, Y., **Tang, T. M.**, and Allen, G. I. [Feature Selection for Data Integration with Mixed Multi-view Data](#). *Annals of Applied Statistics*. 2020.
- Li, X.\*, **Tang, T. M.\***, Wang, X., Kocher, J., and Yu, B. [A stability-driven protocol for drug response interpretable prediction \(staDRIP\)](#). *MLAH: Machine Learning for Health - Extended Abstract (NeurIPS Workshop)*. 2020.
- Altieri, N., Barter, R. L., Duncan, J., Dwivedi, R., Kumbier, K., Li, X., Netzorg, R., Park, B., Singh, C., Tan, Y., **Tang, T. M.**, Wang, Y., Zhang, C., and Yu, B. [Curating a COVID-19 data repository and forecasting county-level death counts in the United States](#). *Harvard Data Science Review*. 2020. [Authors ordered alphabetically]

### Pre-Prints / Under Revision

- Elliott, C. F., Duncan, J.<sup>†</sup>, **Tang, T. M.<sup>†</sup>**, Behr, M., Kumbier, K., and Yu, B. [Designing a Data Science simulation with MERITS: A Primer](#). *arXiv:2403.08971*. 2024.

\* denotes co-first authorship

† denotes equal contribution

Wang, Q.\*, **Tang, T. M.\***, Youlton, N., Weldy, C. S., Kenney, A. M., Ronen, O., Hughes, J. W., Chin, E. T., Sutton, S. C., Agarwal, A., Li, X., Behr, M., Kumbier, K., Moravec, C. S., Tang, W., Margulies, K. B., Cappola, T. P., Butte, A. J., Arnaout, R. A., Brown, J. B., Priest, J. R., Parikh, V. N., Yu, B., and Ashley, E. A. [Epistasis regulates genetic control of cardiac hypertrophy](#). *medRxiv:10.1101/2023.11.06.23297858*. 2024.

Agarwal, A.\*, Kenney, A. M.\*, Tan, Y.\*, **Tang, T. M.\***, and Yu, B. [MDI+: A Flexible Random Forest-Based Feature Importance Framework](#). *arXiv:2307.01932*. 2023.

## INVITED TALKS

<b>Inaugural Berkeley-Stanford Workshop on Veridical Data Science.</b> Low-signal iterative random forests.	2024
<b>University of Minnesota Department of Biostatistics Seminar.</b> Towards reliable experimental recommendations of gene-gene interactions.	2024
<b>University of Michigan Statistics Student Seminar.</b> Interpretable Machine Learning for Reliable Hypothesis Generation.	2024
<b>IMS International Conference on Statistics and Data Science.</b> MDI+: A Flexible Random Forest-Based Feature Importance Framework.	2023
<b>CMStatistics.</b> Integrated Principal Components Analysis.	2023
<b>Joint Statistical Meetings</b> (topic-contributed). MDI+: A Flexible Random Forest-Based Feature Importance Framework.	2023
<b>EcoSta.</b> MDI+: A Flexible Random Forest-Based Feature Importance Framework.	2023
<b>IMS/ASA Spring Research Conference.</b> Integrated Principal Components Analysis.	2023
<b>University of California, Berkeley Statistics Student Seminar.</b> MDI+: A Flexible Random Forest-based Feature Importance Framework.	2023
<b>University of California, Berkeley Biostatistics Joint Group Meeting.</b> simChef culinary school: Cooking up reliable simulations in R.	2023
<b>Bay Area Open Science Group.</b> Curating a COVID-19 data repository and forecasting county-level death counts in the United States.	2021
<b>INFORMS Annual Meeting.</b> A stability-driven protocol for drug response interpretable prediction (staDRIP).	2021
<b>Joint Statistical Meetings</b> (topic-contributed). Curating a COVID-19 data repository and forecasting county-level death counts in the United States.	2021
<b>Women in Data Science Berkeley.</b> Curating a COVID-19 data repository and forecasting county-level death counts in the United States.	2021
<b>Data Science Conference on COVID-19.</b> Curating a COVID-19 data repository and forecasting county-level death counts in the United States: A data perspective.	2020

## CONTRIBUTED TALKS AND POSTERS

<b>Michigan Institute for Data &amp; AI in Society (MIDAS) Future Leaders Summit.</b> Interpretable network-assisted prediction via random forests. (Poster)	2024
<b>Women in Statistics and Data Science.</b> Epistasis regulates genetic control of cardiac hypertrophy.	2022
<b>University of California, Berkeley, Computational Biology Retreat.</b> Detecting genetic and epistatic predictors of increased left ventricular mass. (Poster)	2021
<b>ML4H: Machine Learning for Health (NeurIPS Workshop).</b> A stability-driven protocol for drug response interpretable prediction (staDRIP). (Poster)	2020
<b>Joint Statistical Meetings.</b> Integrated Principal Components Analysis.	2019
<b>WNAR Meeting.</b> Integrated Principal Components Analysis.	2019

## ACADEMIC AND RESEARCH AWARDS

Michigan Institute for Data & AI in Society (MIDAS) Future Leaders Summit	2024
University of California, Berkeley Department of Statistics Evelyn Fix Prize	2023
American Statistical Association SF Bay Area Chapter, Student Travel Award	2023
IMS/ASA Spring Research Conference Student Travel Award	2023
National Science Foundation Graduate Research Fellowship	2019-2023
National Defense Science and Engineering Graduate Fellowship (Declined)	2019-2022
NISS Statistically Accurate Interactive Displays in Graphics Competition, Best Entry Point	2022
Genentech Summer Intern Project Showcase - Product Development Winner	2020
University of California, Berkeley Outstanding Graduate Student Instructor Award	2019-2020
JSM Biometrics Section David P. Byar Young Investigator Award	2019
WNAR Most Outstanding Student Paper Award	2019
Rice Engineering Alumni Senior Statistics Merit Award	2018
J. Venn Leeds Award for Excellence in Scholarship	2018
Rice University Bray Prize in Mathematics	2017
Rice Engineering Alumni Junior Statistics Merit Award	2017
Louis J. Walsh Scholarship in Engineering	2016, 2017
Rice Undergraduate Research Symposium - 1st prize in the Social Sciences division	2016

## SERVICE AND OUTREACH

NISS Graduate Student Network Research Conference, <i>Judge</i>	2024
MiRcore Annual High School Research Conference, <i>Judge</i>	2024
Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS), <i>Judge</i>	2024
UC Berkeley Statistics Graduate Student Association Service Committee, <i>Chair</i>	2019-2023
Expanding Your Horizons Conference, <i>Volunteer</i>	2022
ENVISION Research Competition by WiSTEM, <i>Judge</i>	2022
UCB/UW/UMich Statistics PhD Panel, <i>Panelist</i>	2021
UC Berkeley Statistics NSF GRFP Workshop, <i>Co-organizer</i>	2019-2021
Rice DataSci Club, <i>Co-founder</i>	2017-2018

### *Conference Session Chair/Organizer*

<b>Joint Statistical Meetings</b> , Interpretable machine learning for genomics and biomedical problems, <i>Topic-contributed Session Co-organizer</i>	2024
<b>Inaugural Berkeley-Stanford Veridical Data Science Workshop</b> , <i>Co-organizer</i>	2024
<b>EcoSta</b> , Causal machine learning with high dimensional modeling, <i>Session Chair</i>	2023

## TEACHING

### *University of Notre Dame*

ACMS 40950/80870 Topics in Statistics	Fall 2024
---------------------------------------	-----------

### *University of California, Berkeley (Graduate Student Instructor)*

STAT 154 Modern Statistical Prediction and Machine Learning	Fall 2022
STAT 215A Applied Statistics (Core 1st year PhD graduate course)	Fall 2019

## SOFTWARE

### *R Packages*

<b>simChef</b> ( <i>Author</i> )	For cooking up realistic, reliable, reproducible, and responsible simulations with a tidy grammar
----------------------------------	---

**dgpoix** (*Author*) A modular library of flexible, composable, and real-data-inspired data-generating processes

**vdocs** (*Author*) For beautiful, rigorous, and transparent documentation of PCS-style data analyses

**vthemes** (*Author*) A library of modern plotting themes, color schemes, and R Markdown templates

*Python Packages*

**imodels** (*Contributor*) An sklearn-compatible python package for fitting interpretable models