Erin Craig

ercr@umich.edu | erincraig.me

Biostatistician specializing in interpretable statistical and machine learning methods and software, with expertise in CAR-T therapy, immune repertoire analysis, and precision oncology.

ACADEMIC APPOINTMENTS

Assistant Professor, Department of Biostatistics, University of Michigan 2025-Present

EDUCATION

Ph.D. Biomedical Data Science, Stanford 2019–2025
Advised by: Robert Tibshirani
M.S. Data Science, New College of Florida 2015–2017
B.A. Mathematics, New College of Florida 2005–2009

PUBLICATIONS, SOFTWARE AND PATENTS

PUBLICATIONS

PUBLISHED

1. Liebenberg, **Craig‡**, Tibshirani, Eberlin‡ (‡co-corresponding author.) "Structure-preserving multivariate hypothesis testing for mass spectrometry imaging and single-cell data." Bioinformatics 42.4 (2026): btag137.
2. Wirz, Kotagiri, Haraguchi, Röltgen, Hunter, **Craig**, et al. (2026). Multiplexed antigen panel analysis identifies B cell phenotype and receptor genetic contributions to antibody breadth." *Immunity* 59.5 (2026): 1438-1453.
3. **Craig,** Pilanci, Le Menestrel, Narasimhan, Rivas, Gullaksen, Dehghannasiri, Salzman, Taylor, Tibshirani. Pretraining and the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 88.1 (2026): 261-281.
4. Wirz, Kotagiri, Haraguchi, Röltgen, Hunter, **Craig**, Afaghani et al. "Drivers of differential antibody breadth in SARS-CoV-2 mRNA vaccination and infection." *The Journal of Immunology* 214, no. Supplement_1 (2025): vkaf283-2440.
5. Le Menestrel, **Craig**, Tibshirani, Hastie, Rivas. Using pre-training and interaction modeling for ancestry-specific disease prediction using multiomics data from the UK Biobank. *PLoS One.* 2025;20(12):e0336861. doi:10.1371/journal.pone.0336861.
6. **Craig*,** Keyes* et al. (*co-first author.) Annotation-free discovery of disease-relevant cells in single-cell datasets. *Sci Adv*. 2025;11(35):eadv5019. doi:10.1126/sciadv.adv5019.
7. Zaslavsky*, **Craig*** et al. (*co-first author.) Disease diagnostics using machine learning of B cell and T cell receptor sequences. *Science*. 2025;387(6736):eadp2407. doi:10.1126/science.adp2407.
8. **Craig,** Zhong, Tibshirani. A review of survival stacking: a method to cast survival regression analysis as a classification problem. *Int J Biostat*. 2025;21(1):37–51. doi:10.1515/ijb-2022-0055.

9. Hamilton, **Craig** et al. CAR19 monitoring by peripheral blood immunophenotyping reveals histology-specific expansion and toxicity. *Blood Adv.* 2024;8(12):3314-3326. doi:10.1182/bloodadvances.2024012637.

10. **Craig**, Kroner. Fraud by the Numbers: Medicare Data is a Fraud-Fighting Resource. Taxpayers Against Fraud. September 13, 2023. Accessed February 13, 2026. https://www.taf.org/fbtn2023-sept13/.

11. **Craig,** Poimenidou. An extension of Wolfram's rule 90 for one-dimensional cellular automata over non-abelian group alphabets. *A Celebration of the EDGE Program's Impact on the Mathematics Community and Beyond*. 2019: 289-316.

12. Sherman, **Craig**, Yanovich E., Ketko, Kalmanovich, Yanovich R. Standardized interpretation of heat-tolerance-testing results: probability of intolerance instead of specialist judgment. *Journal of Athletic Training*. 2018; 53(4): 423-430.

13. **Craig**, Arias, Gillman. Predicting readmission risk from doctors' notes. *NeurIPS Workshop on Machine Learning for Healthcare*. 2017.

14. Childers*, **Craig***, Taha, Poimenidou. (*co-first author.) A p-adic approach to binomial identities. *Pi Mu Epsilon Journal*. 2010: 133-142.

IN PREPARATION AND UNDER REVIEW

1. **Craig**, Tibshirani. Supervised learning pays attention. *arXiv preprint arXiv:2512.09912* (2026).

2. Gullaksen, Brevik, Sefland, **Craig**, et al. Deep single-cell immune and signaling profiles predict long term therapy response in chronic myeloid leukemia within hours. *bioRxiv preprint: https://doi.org/10.1101/2025.11.10.687542*. 2025-11 (2025).

SOFTWARE

1. Keyes*, **Craig***. (**co-first author*). Open source R package. MMIL: Mixture Models for Multiple Instance Learning. In beta testing.

2. **Craig**, Tibshirani. Open source R package: "ptLasso: Pretraining for the Lasso." GitHub repository, erincr.github.io/ptLasso/ (2024).

3. **Craig**, Tibshirani. Open source R package: "sweetspot: Find and Assess Treatment Effect Sweet Spots in Clinical Trials." GitHub repository, https://github.com/erincr/sweetspot (2020).

PATENTS

1. Zaslavsky, **Craig**, Michuda, Sehgal, Tibshirani, Boyd. Systems and Methods for Assessment of Immune Response and Applications Thereof. 63/553,078, filed 2/13/2024. Patent pending.

PRESENTATIONS

UPCOMING INVITED TALKS

1. *Pretraining and the lasso*. JRSS Series B Editors' Invited Session, Royal Statistical Society International Conference. Bournemouth, England. September 2026.

INVITED TALKS (EXTRAMURAL)

1. Machine learning for immune repertoire data. The Fries Lab, University of Rochester. Rochester, NY. [Talk given remotely on Zoom.] March 2026.
2. *An introduction to biostatistics*. University of Oklahoma College of Medicine. Oklahoma City, OK. [Talk given remotely on Zoom.] February 2026.
3. *Pretraining and the lasso*. Joint Statistical Meetings, Session: Harnessing the power of large-scale and heterogeneous data with integrative analysis. Nashville, TN, August 2025.
4. *Pretraining and the lasso*. Women in Data Science: Health Frontiers. Stanford University, Stanford, CA. May 2024.
5. *Treatment effect sweet spots in clinical trials*. Stanford CORES Conference. Stanford University, Stanford, CA. May 2023.
6. *Stacking: Survival Analysis as a Classification Problem*. AAAI Survival Prediction Symposium. Remote. March 2021.
7. *Predicting Hospital Readmissions from Doctors' Notes*. Hofstra Mathematics Department. Remote. December 2020.

CONFERENCE TALKS

8. Plenary talk: *Using the lasso to identify disease-specific B cell receptors*. NLM T15 Training Conference, Stanford University, Stanford, CA, June 2023.
   Earned best talk across 18 biomedical training programs in the US.
9. *Treatment effect sweet spots in clinical trials*. Stanford Causal Science Conference, Stanford University, Stanford, CA, June 2023.

INTERNAL TALKS AND SEMINARS

1. *Supervised learning pays attention*. AI in Biostatistics Lunch Discussion, Department of Biostatistics. University of Michigan, Ann Arbor, MI. February 2026.
2. *Interpretable prediction for high-dimensional, heterogeneous cancer data*. Cancer Genomics and Signaling Tumor Microenvironment Seminar, Rogel Cancer Center. University of Michigan, Ann Arbor, MI. January 2026.
3. *Identifying disease-associated cells in single-cell datasets*. Guest lecture, BIOSTAT 803: Biostatistics in Cancer Seminar. University of Michigan, Ann Arbor, MI. December 2025.
4. *Interpretable learning methods for biomedical data: a "choose your own adventure" talk*. e-HAIL Seminar Series (E-Health and Artificial Intelligence). University of Michigan, Ann Arbor, MI. November 2025.
5. *Interpretable learning methods for biomedical data: a "choose your own adventure" talk*. Biostatistics seminar. University of Michigan, Ann Arbor, MI. September 2025.

TEACHING

University of Michigan, School of Public Health

Co-instructor, Case Studies in Health Big Data (BIOSTAT 629), 2026

Stanford University

T.A., Biomedical Informatics Research Methodology (BIOMEDIN 212), 2022
T.A., Data Driven Medicine (BIOMEDIN 215), 2020

New College of Florida

Instructor, Data Science Seminar, 2015-2016
T.A., Algorithms and Optimization (graduate), 2015
T.A., Linear Algebra (undergraduate), 2007
T.A., Calculus (undergraduate), 2006

Wolfram Summer Schools, 2010–2014

Wrote programming curriculum and mentored advanced high school students through projects. Taught cellular automata and programming in Mathematica.

Additional experience

Ballet teacher (beginner through professional), 2001–2015
Fitness teacher (barre and TRX), 2013–2015

GRANTS

CURRENT

MIDAS DATA Fund (NSF IUCRC Pilot Grant)

Type: Pilot funding
Sponsor: Center for Data-Driven Drug Development and Treatment Assessment (DATA IUCRC) and Michigan Institute for Data & AI in Society (University of Michigan).
Role: PI; co-I Dr. Maor Sauler (Michigan Medicine, Pulmonary)
Dates: 1 year, beginning summer 2026
Total award amount: $40,000
Status: Awarded
Project: Pathway-Informed Cellular and Patient Phenotypes from Single-Cell Transcriptomics

e-HAIL Summer Student Support

Type: Student summer funding
Sponsor: University of Michigan, e-HAIL
Role: Co-PI, with Dr. Fatima Fattahi (Michigan Medicine, Internal Medicine)
Dates: Summer 2026
Total award amount: $15,000
Status: Awarded
Project: Interpretable Patient-Level Modeling of Type 2–High and Type 2–Low Asthma Using Single-Cell Transcriptomics

PREVIOUS

Stanford High Impact Technology Fund
Stanford University High Impact Technology Fund Award, awarded for the commercialization of an innovative precision diagnostics platform for autoimmune disease, leading to the formation of a startup that has raised $8M.

Type: Seed fund
Sponsor: Stanford University, Office of Technology Licensing (High Impact Technology Fund)
Role: Co-investigator (methods development)
Dates: 2023 (1 year; 4 quarters)
Total award amount: $22,000, industry advisory support, and an MBA fellow
Status: Completed

Stanford Data Science Scholarship
Stipend and half tuition awarded to fewer than 20 graduate students per year.

Type: Fellowship
Sponsor: Stanford University, Vice Provost for Graduate Education (VPGE)
Role: Fellow
Dates: 09/2022–08/2024 (2 years; 8 quarters)
Total award amount: Tuition and $45,000/year stipend for three years.
Status: Completed

Stanford Graduate Fellowship (Gabilan Fellow)
Stipend and tuition awarded to fewer than 100 graduate students per year.

Type: Fellowship
Sponsor: Stanford University, Vice Provost for Graduate Education (VPGE)
Role: Fellow
Dates: 09/2019–08/2022 (3 years; 12 quarters)
Total award amount: Tuition and $45,000/year stipend for three years.
Status: Completed

AWARDS AND HONORS

Gold medal: best plenary talk at NLM T15 Training Conference, 2023

Earned best talk across 18 biomedical training programs in the US; talk title "Using the lasso to identify disease-specific B cell receptors".

PROFESSIONAL SERVICE

1. Member, Student Recruitment Committee, Biostatistics, University of Michigan. AY 2025-2026. Gave recruitment outreach talks for the following student groups:
    - Girls in Electrical Engineering and Computer Science. November 2025.
    - Michigan Data Science Team. October 2025.
    - Michigan Undergraduate Students of Statistics. October 2025.
    - Michigan Undergraduate Honors Program. September 2025.

2. Member, Internal Advisory Board, AI & Digital Health Innovation, University of Michigan, 2026–present.
3. Co-organizer, Pulmonology and Biostatistics Monthly Seminar, University of Michigan, 2026-present.
4. Organizer, Stanford Data Science Conference, Stanford University. May 2023 and May 2024.
5. Lead organizer, Women in Data Science Conference, New College of Florida, February 2017.
6. Organizer, New College of FL, Data Science Seminar, AY 2015-2016.
7. Organizer, DataKind "Data Dive", San Francisco CA, August 2016.

JOURNAL REVIEWS

Ad-hoc reviewer, manuscript reviews:

1. PeerJ, 2024-present (2 reviews).
2. International Journal of Biostatistics, 2026-present (1 review).
3. Annals of Applied Statistics. 2026-present (1 review).
4. Communications in Statistics. 2026-present (1 review).
5. Mathematical Biosciences and Engineering. 2026-present (1 review).

GRANT REVIEWS

1. Propelling Original Data Science (PODS) pilot grants, Michigan Institute for Data & AI in Society (MIDAS), University of Michigan. 2026 (3 reviews).

INDUSTRY RESEARCH AND SOFTWARE EXPERIENCE

Jonathan Kroner Law Office: *Data Scientist* 2015–Present
Data-driven whistleblower. Support fraud investigations by analyzing Medicare data. (~5 hours annually)

BlackRock AI Labs: *Data Science Intern* Summer 2021
Developed predictive models to inform trading decisions.

Florence A. Rothman Institute: *Data Scientist* 2017–2018
Used deep learning on text to build predictive models for hospitals.

Wolfram|Alpha: *Manager/Lead Developer* 2009–2015, 2019
Architected and led development of Step-by-step Solutions and the Wolfram Problem Generator. Oversaw math content development. Led teams to build and maintain education technologies.

EXTRACURRICULAR ACHIEVEMENTS

Contemporary Dancer 2009-2017
Performed professionally in San Francisco CA, Sarasota FL, Boston MA and Cleveland OH. Studied at the Joffrey Ballet, the San Francisco Conservatory of Dance, the Martha Graham Winter Intensive and Ballet Chicago; performed works by William Forsythe, Jiří Kylián, Ohad Naharin and George Balanchine.