Jaron Arbet
SENIOR BIOINFORMATICIAN
UCLA Jonsson Comprehensive Cancer Center

✉ jaron.arbet@gmail.com | 🏠 jarbet.github.io | 🎓 Google Scholar | ⓞ Github

📖 Research Interests

- Machine learning and predictive modeling
- High-dimensional "Big Data"
- Variable selection
- Multi-omics cancer data integration
- Robust nonparametric statistics

💼 Work Experience

Senior Bioinformatician   *April 2022 - Present*

JONSSON COMPREHENSIVE CANCER CENTER, UNIVERSITY OF CALIFORNIA LOS ANGELES, CA

Staff Biostatistician   *Jan 2021 - April 2022*

EDWARDS LIFESCIENCES, IRVINE, CA

- Develop statistical analysis plans and data monitoring for clinical trials
- Power and sample size estimation
- Clinical study reports for federal regulatory agencies
- Perform analyses with large national database (STS/ACC TVT Registry) including propensity score matching for causal inference

Postdoctoral Research Associate   *Aug 2018 - Jan 2021*

UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS, DEPARTMENT OF BIOSTATISTICS AND INFORMATICS, AURORA, CO

- Provided statistical consulting services
- Published scientific manuscripts
- Developed grants (statistical analysis plan, power/sample size estimation)
- Supervised students and MS/MPH degree biostatisticians
- Taught BIOS 6601 Intro stats course for 75 MPH students, and created annual 6-week data science machine learning short course
- Develop new statistical methodology

Research Assitant   *June 2014 - July 2018*

UNIVERSITY OF MINNESOTA, DIVISION OF BIOSTATISTICS, MINNEAPOLIS, MN

- Developed variable selection methods for high dimensional genetic data and robust methods for estimating the genetic heritability of disease traits

Statistical Genetics Research Intern   *June - Aug 2013*

DORDT COLLEGE, DEPARTMENT OF MATHEMATICS, SIOUX CENTER, IA

- Developed methods for aggregating different statistical tests together for greater sensitivity in detecting disease-related genes; resampling-based bias corrected tests

Math and Statistics Tutor   *Sept 2012 - May 2014*

WINONA STATE UNIVERSITY, WINONA, MN

🎓 Education

University of Minnesota   *Minneapolis, MN, USA*

PHD BIOSTATISTICS   *2014-2018*

- GPA: 3.7

Wiona State University   *Winona, MN, USA*

B.S. MATHEMATICS AND STATISTICS   *2010-2014*

- GPA: 4.0 | *Summa Cum Laude*

📕 Peer-reviewed Publications

1. Lippitt, Carlson, **Arbet,** Fingerlin, Maier, & Kechris. (2024). Limitations of clustering with PCA and correlated noise. *Journal of Statistical Computation and Simulation*. `https://doi.org/10.1080/00949655.2024.2329976`

2. Ni, Rogowitz, Farahmand, Kaizer, **Arbet,** Cunningham, Thomas, & Saxon. (2024). Weight loss outcomes in a veterans affairs pharmacotherapy-based weight management clinic. *Journal of the Endocrine Society*. `https://doi.org/10.1210/jendso/bvae042`

3. Tohi, Sahrmann, **Arbet,** Kato, Lee, Peacock, Ginsburg, Pavlovich, Carroll, Bangma, Sugimoto, & Boutros. (2024). De-escalation of monitoring in active surveillance for prostate cancer: Results from the GAP3 consortium. *European Urology Oncology*. `https://doi.org/10.1016/j.euo.2024.07.006`

4. Weiner, Agrawal, Wang, Sonni, Li, **Arbet,** Zhang, Proudfoot, Hong, Davicioni, Kane, Valle, Kishan, Pra, Ghadjar, Sweeney, Nickols, Karnes, Shen, Rettig, Czernin, Ross, Chua, L. K., Schaeffer, Calais, Boutros, & Reiter. (2024). Molecular hallmarks of prostate-specific membrane antigen in treatment-na<u+00EF>ve prostate cancer. *European Urology*. `https://doi.org/10.1016/j.eururo.2024.09.005`

5. Chan, Dodson, **Arbet,** Boutros, & Xiao. (2023). Single-cell analysis in lung adenocarcinoma implicates RNA editing in cancer innate immunity and patient prognosis. *Cancer Research*. `https://doi.org/10.1158/0008-5472.CAN-22-1062`

6. Gamallat, Choudhry, Li, Rokne, Alhajj, Abdelsalam, Ghosh, **Arbet,** Boutros, & Bismar. (2023). Serrate RNA effector molecule (SRRT) is associated with prostate cancer progression and is a predictor of poor prognosis in lethal prostate cancer. *Cancers*. `https://doi.org/10.3390/cancers15102867`

7. Greca, Grau, **Arbet,** Liao, Sosa, Haugen, & Kitahara. (2023). Anthropometric, dietary, and lifestyle factors and risk of advanced thyroid cancer: The NIH-AARP diet and health cohort study. *Clinical Endocrinology*. `https://doi.org/10.1111/cen.14970`

8. Creasy, Ostendorf, Blankenship, Grau, **Arbet,** Bessesen, Melanson, & Catenacci. (2022). Effect of sleep on weight loss and adherence to diet and physical activity recommendations during an 18-month behavioral weight loss intervention. *International Journal of Obesity (2005)*. `https://doi.org/10.1038/s41366-022-01141-z`

9. Grau, **Arbet,** Ostendorf, Blankenship, Panter, Catenacci, Melanson, & Creasy. (2022). Creating an algorithm to identify indices of sleep quantity and quality from a wearable armband in adults. *Sleep Science (Sao Paulo, Brazil)*. `https://doi.org/10.5935/1984-0063.20220052`

10. Lin, **Arbet,** Mroz, Liao, Restrepo, Mayer, Li, Barkes, Schrock, Hamzeh, Fingerlin, Carlson, & Maier. (2022). Clinical phenotyping in sarcoidosis using cluster analysis. *Respiratory Research*. `https://doi.org/10.1186/s12931-022-01993-z`

11. Okamoto, Devoe, Seto, Minarchick, Wilson, Rothfuss, Mohning, **Arbet,** Kroehl, Visser, August, Thomas, Charry, Fleischer, Feser, Frazer-Abel, Norris, Cherrington, Janssen, Kaplan, Deane, Holers, & Demoruelle. (2022). Association of sputum neutrophil extracellular trap subsets with IgA anti-citrullinated protein antibodies in subjects at risk for rheumatoid arthritis. *Arthritis & Rheumatology (Hoboken, N.J.)*. `https://doi.org/10.1002/art.41948`

12. Wood, **Arbet,** Amura, Nodine, Collins, Orlando, Mayer, Stein, & Anderson. (2022). Multicenter study evaluating nitrous oxide use for labor analgesia at high- and low-altitude institutions. *Anesthesia and Analgesia*. `https://doi.org/10.1213/ANE.0000000000005712`

13. **Arbet,** Zhuang, Litkowski, Saba, & Kechris. (2021). Comparing statistical tests for differential network analysis of gene modules. *Frontiers in Genetics*. `https://doi.org/10.3389/fgene.2021.630215`

14. Carpenter, Frank, Williamson, **Arbet,** Wagner, Kechris, & Kroehl. (2021). tidyMicro: A pipeline for microbiome data analysis and visualization using the tidyverse in r. *BMC Bioinformatics*. `https://doi.org/10.1186/s12859-021-03967-2`

15. Nodine, **Arbet,** Jenkins, Rosenthal, Carrington, Purcell, Lee, & Hoon. (2021). Graduate nursing student stressors during the COVID-19 pandemic. *Journal of Professional Nursing : Official Journal of the American Association of Colleges of Nursing*. `https://doi.org/10.1016/j.profnurs.2021.04.008`

16. Ostendorf, Blankenship, Grau, **Arbet,** Mitchell, Creasy, Caldwell, Melanson, Phelan, Bessesen, & Catenacci. (2021). Predictors of long-term weight loss trajectories during a behavioral weight loss intervention: An exploratory analysis. *Obesity Science & Practice*. `https://doi.org/10.1002/osp4.530`

17. Ramakrishnan, **Arbet,** Mace, Suresh, Smith, S., Soler, & Smith. (2021). Predicting olfactory loss in chronic rhinosinusitis using machine learning. *Chemical Senses*. `https://doi.org/10.1093/chemse/bjab042`

18. Reed, **Arbet,** & Staubli. (2021). Clinical nurse specialists in the united states registered with a national provider identifier. *Clinical Nurse Specialist CNS*. `https://doi.org/10.1097/NUR.0000000000000592`

19. Rosenthal, Lee, Jenkins, **Arbet,** Carrington, Hoon, Purcell, & Nodine. (2021). A survey of mental health in graduate nursing students during the COVID-19 pandemic. *Nurse Educator*. `https://doi.org/10.1097/NNE.`

0000000000001013

20. Schmanski, Roberts, Coors, Wicks, **Arbet,** Weber, Crooks, Barnes, & Taylor. (2021). Research participant understanding and engagement in an institutional, self-consent biobank model. *Journal of Genetic Counseling*. `https://doi.org/10.1002/jgc4.1316`

21. **Arbet,** Brokamp, Meinzen-Derr, Trinkley, & Spratt. (2020). Lessons and tips for designing a machine learning study using EHR data. *Journal of Clinical and Translational Science*. `https://doi.org/10.1017/cts.2020.513`

22. **Arbet,** McGue, & Basu. (2020). A robust and unified framework for estimating heritability in twin studies using generalized estimating equations. *Statistics in Medicine*. `https://doi.org/10.1002/sim.8564`

23. Coleman-Minahan, Sheeder, **Arbet,** & McLemore. (2020). Interest in medication and aspiration abortion training among colorado nurse practitioners, nurse midwives, and physician assistants. *Women's Health Issues : Official Publication of the Jacobs Institute of Women's Health*. `https://doi.org/10.1016/j.whi.2020.02.001`

24. Gance-Cleveland, Linton, **Arbet,** Stiller, & Sylvain. (2020). Predictors of overweight and obesity in childhood cancer survivors. *Journal of Pediatric Oncology Nursing : Official Journal of the Association of Pediatric Oncology Nurses*. `https://doi.org/10.1177/1043454219897102`

25. Thomas, Zaman, Cornier, Catenacci, Tussey, Grau, **Arbet,** Broussard, & Rynders. (2020). Later meal and sleep timing predicts higher percent body fat. *Nutrients*. `https://doi.org/10.3390/nu13010073`

26. James-Allan, **Arbet,** Teal, Powell, & Jansson. (2019). Insulin stimulates GLUT4 trafficking to the syncytiotrophoblast basal plasma membrane in the human placenta. *The Journal of Clinical Endocrinology and Metabolism*. `https://doi.org/10.1210/jc.2018-02778`

27. **Arbet,** McGue, Chatterjee, & Basu. (2017). Resampling-based tests for lasso in genome-wide association studies. *BMC Genetics*. `https://doi.org/10.1186/s12863-017-0533-3`

28. Grinde, **Arbet,** Green, O'Connell, Valcarcel, Westra, & Tintle. (2017). Illustrating, quantifying, and correcting for bias in post-hoc analysis of gene-based rare variant tests of association. *Frontiers in Genetics*. `https://doi.org/10.3389/fgene.2017.00117`

29. Greco, Hainline, **Arbet,** Grinde, Benitez, & Tintle. (2016). A general approach for combining diverse rare variant association tests provides improved robustness across a wider range of genetic architectures. *European Journal of Human Genetics : EJHG*. `https://doi.org/10.1038/ejhg.2015.194`

⚕ Awards

POSTDOCTORAL

Staff appreciation and recognition (STAR) *2023-2024*
UCLA JONSSON COMPREHENSIVE CANCER CENTER

Staff appreciation and recognition (STAR) *2022-2023*
UCLA JONSSON COMPREHENSIVE CANCER CENTER

STUDENT AWARDS

Outstanding Research Assistant Award *2017-2018*
UNIVERSITY OF MINNESOTA, DIVISION OF BIOSTATISTICS & HEALTH DATA SCIENCE

Interdisciplinary Biostatistics Training Grant in Genetics and Genomics *2015-2017*
NATIONAL INSTITUTE OF GENERAL MEDICAL SCIENCES

COMPETITIONS

Midwest Undergraduate Data Analysis Competition (MUDAC) *2014*
SCHOOL OF EDUCATION, UNIVERSITY OF MICHIGAN
• 3rd place out of 20 teams from 14 universities

1st place in the "predictive-modeling class competition" *2014*
WINONA STATE UNIVERSITY DATA MINING COURSE

Midwest Undergraduate Data Analysis Competition (MUDAC) *2013*
SCHOOL OF EDUCATION, UNIVERSITY OF MICHIGAN
• 2nd place out of 15 teams from 10 universities

📊 Presentations

◼ Departmental

Robust regression for noisy biological data
Statistics Working Group, Cancer Data Sciences, University of California, Los Angeles, CA *2024*
- https://jarbet.github.io/presentation-ordinal-regression/ordinal_regression.html

Linear regression diagnostics
Boutros Lab, Cancer Data Sciences, University of California, Los Angeles, CA *2024*
- https://jarbet.github.io/presentation-regression-diagnostics/regression_diagnostics.html

Causal inference with observational data using propensity score matching
Statistics Working Group, Cancer Data Sciences, University of California, Los Angeles, CA *2024*
- https://jarbet.github.io/presentation-causal-inference-matching/causal_inference_matching.html

Automatic feature selection and engineering with MARS
Statistics Working Group, Cancer Data Sciences, University of California, Los Angeles, CA *2023*
- https://jarbet.github.io/presentation-mars.github.io/presentation-mars.html

Introduction to Bayesian statistics
Boutros Lab, Cancer Data Sciences, University of California, Los Angeles, CA *2022*
- https://jarbet.github.io/presentations/2022-08-25_Bayesian_stats.pdf

Multiomics cancer data analysis
Boutros Lab, Cancer Data Sciences, University of California, Los Angeles, CA *2021*
- https://jarbet.github.io/presentations/2021-11-21_multiomics_cancer.pdf

Interpretable machine learning
Center for Innovative Design and Analysis, University of Colorado, Denver, CO *2021*
- https://jarbet.github.io/presentations/2020-01-21_interpretable_machine_learning.pdf

Differential Network Analysis of Gene Modules
Systems Genetics and Bioinformatics Research Lab, University of Colorado, Denver, CO *2021*

Frequentist versus Bayesian Approaches to Statistics: A Debate
Department of Biostatistics and Informatics, University of Colorado, Denver, CO *2018*

Robust Estimation of Genetic and Environmental Variance Components in Twin Studies
University of Minnesota Biostatistics Student Seminar Series, Minneapolis, MN *2017*

A General framework for Association Tests with Multivariate Traits in Large-Scale Genomics Studies
University of Minnesota Statistical Genetics Journal Club, Minneapolis, MN *2016*

Methods of Inference for Penalized Regression in High-Dimensional Genetic Association Studies
School of Public Health Research Day, University of Minnesota, Minneapolis, MN *2016*

Penalized Regression for High-Dimensional Genetic Association Testing
University of Minnesota Statistical Genetics Journal Club, Minneapolis, MN *2016*

◼ Conferences

Methylome-Genome Interactions Define Disease Aggression in Localized Prostate Cancer
Poster
RECOMB/ISCB Conference on Regulatory and Systems Genomics, Los Angeles, CA *2024*

Triplet matching: propensity score matching with 3 groups
Talk
14th Annual FDA/AdvaMed Medical Device Statistical Issues Conference (Virtual) *2022*
- https://jarbet.github.io/presentations/20220422_triplet_matching.pdf

Statistical Considerations and Methods to Utilize Real World Evidence in Medical Device Evaluation
Panel Co-organizer
14th Annual FDA/AdvaMed Medical Device Statistical Issues Conference (Virtual) *2022*

Robust Estimation of Genetic and Environmental Variance Components in Twin Studies using Generalized Estimating Equations

Poster

JOINT STATISTICAL MEETINGS (JSM), BALTIMORE, MD

2017

Methods of Inference for Penalized Regression in High-Dimensional Genetic Association Studies

Poster

EASTERN NORTH AMERICAN REGION INTERNATIONAL BIOMETRIC SOCIETY (ENAR), AUSTIN, TX

2016

What now? Post-Hoc Approaches for Gene-based Rare-Variant Tests of Association

Poster

AMERICAN SOCIETY OF HUMAN GENETICS CONFERENCE (ASHG), BOSTON, MA

2013

General Approaches for Combining Rare Variant Association Tests Provide Improved Power Across a Wide Range of Genetic Architectures

Poster

AMERICAN SOCIETY OF HUMAN GENETICS CONFERENCE (ASHG), BOSTON, MA

2013

👨‍🏫 Teaching Experience

▪ SEMESTER COURSES

BIOS 6601 – Applied Biostatistics I

Fall 2019

UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

- *Duty*: Sole Instructor, completely redesigned all lectures, homework, quizzes, and exams
- *Description*: Applied biostatistical methods including descriptive and statistical inference; odds ratio and relative risk, probability theory, parameter estimation, tests for comparing statistics of two or more groups, correlation, linear regression, logistic regression and survival analysis.
- *Credits*: 3
- *Audience*: 75 Masters of Public Health (MPH) students

▪ SHORT COURSES

Fundamentals of Data Science Literacy

Summer 2019, Fall 2020

UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

- *Duty*: lead organizer, developed and taught 3 machine learning lectures
- *Description*: 6-week short course (2 hours/week) on data science with emphasis on how machine learning and data mining are used in biomedical research. Topics included: 1) Overview of Data Science; 2) Introduction of Machine Learning in Biomedical Research; 3) Clustering and Pattern Finding (unsupervised Machine Learning); 4) Powerful Predictive Models and Variable Selection (supervised Machine Learning); 5) Fundamentals of Data Visualization; 6): Reproducible Research and Team Science
- *Audience*: 20 medical researchers, mostly non-statisticians

Introduction to Statistics

Summer 2019, 2020

UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

- *Duty*: Instructor, developed and taught lectures on "Introduction to statistics", study design, and survey design
- *Description*: Introductory statistics day-long short course for UCHealth nurses
- *Audience*: 15 UCHealth Nurses

▪ GUEST LECTURES

IPHY 5900 – Data Literacy in Biomedical Research

Nov 2019, Oct 2020

UNIVERSITY OF COLORADO BOULDER

- *Primary Instructor*: Dr. Celine Marie Claire D. Vetter, Integrative Physiology
- *Lecture*: Machine Learning in Biomedical Research

BIOS 6611 – Biostatistical Methods I

Fall 2018

UNIVERSITY OF COLORADO SCHOOL OF PUBLIC HEALTH

- *Primary Instructor*: Dr. Alex Kaizer
- *Lecture*: Linear Algebra Review

👥 Mentoring and supervising

⬛ UNIVERSITY OF CALIFORNIA LOS ANGELES

Supervising students for data analysis projects

- Yuxi Song
- Sattwik Banerjee
- Adriana Wiggins

⬛ UNIVERSITY OF COLORADO, DEPARTMENT OF BIOSTATISTICS AND INFORMATICS

Supervising MPH/MS Statisticians

- Laura Grau (MPH), Nov 2018 – Jan 2021
  - I supervised Laura on several projects which led to publications.

Supervising students for data analysis projects

- Yaxu Zhuang (PhD), Jan 2019 – Aug 2020
- Rachel Weber (MS), Dec 2018 – April 2019

Thesis or Dissertation Committee Member (graduating year)

- Charlie Carpenter, Biostatistics MS thesis committee, Biostatistics (2020)
- Rachel Blumhagen, Biostatistics PhD dissertation committee (2021)

👉 Professional Service

⬛ GRANT AND PROTOCOL REVIEWS

- Scientific Advisory & Review Committee (SARC), Colorado Clinical and Translational Sciences Institute, University of Colorado – April 2019 to Jan 2021

⬛ JOURNAL REVIEWER

- Biometrics
- Genetic Epidemiology
- PeerJ

⬛ COMMITTEES/OTHER

Department of Biostatistics and Informatics, University of Colorado

- Student Admissions Committee, Fall 2018 – Spring 2019
- Volunteer Judge for CIDA Hackathon (24 hour data analysis competition), Jan. 2020

Biostatistics, Epidemiology, & Research Design Core of the Colorado Clinical and Translational Sciences Institute, University of Colorado

- Education Committee, Fall 2018 – Jan 2020

🖥 Software Skills

Proficient

®

- Over **13 years** of experience with R for data analysis
- I enjoy **developing R packages** to easily share documented code with colleagues and the broader research community. Examples:
  - `https://github.com/arbet003/discoMod`
- **Parallel programming** in R using multiple cpus

🐙 git/github

- Version control
- Used daily at UCLA
- Collaborating with others on projects
- Review colleagues' code
- Assign issues to team members

Quarto and Rmarkdown

- Reproducible research, manuscripts and reports
- Reproducible HTML presentations. Examples: `https://jarbet.github.io/presentations.html`
- Websites: `https://jarbet.github.io/`
- Resume: this resume was made using Rmarkdown with the vitae R package

Bayesian statistical programming

- Proficient with BUGS/JAGS; basic skills with STAN

Basic skills

- Linux command line and bash scripts
  - Daily use of a high-performance computing cluster at UCLA
- Databricks
- Microsoft Office
- SAS