Jaron Arbet, Ph.D.

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

- Analysis of cancer molecular data: whole genome sequencing, RNA-Seq, DNA methylation, proteomics
- Biomarker identification: among thousands of molecular features, which are most associated with cancer outcomes?
- Daily git/github use for version control and collaboration
- Supervised machine learning for predicting cancer prognosis
- Unsupervised machine learning for deriving cancer subtypes
- Supervising undegraduate and graduate students on projects
- Develop open-source R packages for sharing software with broader scientific community
- Create statistical analysis plans and power/sample size estimation for grants

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 Assistant

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. Jones, L. W., Lavery, J. A., Tsai, B. L., Moskowitz, C. S., Lee, C. P., Harrison, J., Michalski, M. G., Stoeckel, K., Graham, C., Iyengar, N. M., Bhanot, U., Linkov, I., Jain, M., Jochelson, M. S., Monetti, M., Seewaldt, V. L., Pilewskie, M. L., Pribil, P., Zhu, C., **Arbet, J.**, Mangino, D. A., & Boutros, P. C. (2025). A co-clinical trial of exercise therapy in breast cancer prevention. *Clinical Cancer Research*, OF1–OF11. https://doi.org/10.1158/1078-0432.ccr-24-4298
2. Tohi, Y., Sahrmann, J. M., **Arbet, J.**, Kato, T., Lee, L. S., Peacock, M., Ginsburg, K., Pavlovich, C., Carroll, P., Bangma, C. H., Sugimoto, M., & Boutros, P. C. (2025). De-escalation of monitoring in active surveillance for prostate cancer: Results from the GAP3 consortium. *European Urology Oncology*, *8*(2), 347–354. https://doi.org/10.1016/j.euo.2024.07.006
3. Xu, X., Zhu, H., Hugh-White, R., Livingstone, J., Eng, S., Zeltser, N., Wang, Y., Pajdzik, K., Chen, S., Houlahan, K. E., Luo, W., Liu, S., Xu, X., Sheng, M., Guo, W. Y., **Arbet, J.**, Song, Y., Wang, M., Zeng, Y., Wang, S., Zhu, G., Gao, T., Chen, W., Ci, X., Xu, W., Xu, K., Orain, M., Picard, V., Hovington, H., Bergeron, A., … He, H. H. (2025). The landscape of N6-methyladenosine in localized primary prostate cancer. *Nature Genetics*, *57*(4), 934–948. https://doi.org/10.1038/s41588-025-02128-y
4. Weiner, A. B., Agrawal, R., Wang, N. K., Sonni, I., Li, E. V., **Arbet, J.**, Zhang, J. J. H., Proudfoot, J. A., Hong, B. H., Davicioni, E., Kane, N., Valle, L. F., Kishan, A. U., Pra, A. D., Ghadjar, P., Sweeney, C. J., Nickols, N. G., Karnes, R. J., Shen, J., Rettig, M. B., Czernin, J., Ross, A. E., Lee Kiang Chua, M., Schaeffer, E. M., Calais, J., Boutros, P. C., & Reiter, R. E. (2024). Molecular hallmarks of prostate-specific membrane antigen in treatment-naïve prostate cancer. *European Urology*, *86*(6), 579–587. https://doi.org/10.1016/j.eururo.2024.09.005
5. Lippitt, W., Carlson, N. E., **Arbet, J.**, Fingerlin, T. E., Maier, L. A., & Kechris, K. (2024). Limitations of clustering with PCA and correlated noise. *Journal of Statistical Computation and Simulation*, *94*(10), 2291–2319. https://doi.org/10.1080/00949655.2024.2329976
6. Ni, K., Rogowitz, E., Farahmand, A. K., Kaizer, L. K., **Arbet, J.**, Cunningham, C. R., Thomas, E. A., & Saxon, D. R. (2024). Weight loss outcomes in a veterans affairs pharmacotherapy-based weight management clinic. *Journal of the Endocrine Society*, *8*(5). https://doi.org/10.1210/jendso/bvae042
7. Greca, A. L., Grau, L., **Arbet, J.**, Liao, L. M., Sosa, J. A., Haugen, B. R., & Kitahara, C. M. (2023). Anthropometric, dietary, and lifestyle factors and risk of advanced thyroid cancer: The NIH-AARP diet and health cohort study. *Clinical Endocrinology*, *99*(6), 586–597. https://doi.org/10.1111/cen.14970
8. Gamallat, Y., Choudhry, M., Li, Q., Rokne, J. G., Alhajj, R., Abdelsalam, R., Ghosh, S., **Arbet, J.**, Boutros, P. C., & Bismar, T. A. (2023). Serrate RNA effector molecule (SRRT) is associated with prostate cancer progression and is a predictor of poor prognosis in lethal prostate cancer. *Cancers*, *15*(10), 2867. https://doi.org/10.3390/cancers15102867
9. Chan, T. W., Dodson, J. P., **Arbet, J.**, Boutros, P. C., & Xiao, X. (2022). Single-cell analysis in lung adenocarcinoma implicates RNA editing in cancer innate immunity and patient prognosis. *Cancer Research*, *83*(3), 374–385. https://doi.org/10.1158/0008-5472.can-22-1062
10. Grau, L., **Arbet, J.**, Ostendorf, D. M., Blankenship, J. M., Panter, S. L., Catenacci, V. A., Melanson, E. L., & Creasy, S. A. (2022). Creating an algorithm to identify indices of sleep quantity and quality from a wearable armband in adults. *Sleep Science*, *15*(03), 279–287. https://doi.org/10.5935/1984-0063.20220052
11. Creasy, S. A., Ostendorf, D. M., Blankenship, J. M., Grau, L., **Arbet, J.**, Bessesen, D. H., Melanson, E. L., & Catenacci, V. A. (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*, *46*(8), 1510–1517. https://doi.org/10.1038/s41366-022-01141-z
12. Lin, N. W., **Arbet, J.**, Mroz, M. M., Liao, S.-Y., Restrepo, C. I., Mayer, A. S., Li, L., Barkes, B. Q., Schrock, S., Hamzeh, N., Fingerlin, T. E., Carlson, N. E., & Maier, L. A. (2022). Clinical phenotyping in sarcoidosis using cluster analysis. *Respiratory Research*, *23*(1). https://doi.org/10.1186/s12931-022-01993-z
13. Okamoto, Y., Devoe, S., Seto, N., Minarchick, V., Wilson, T., Rothfuss, H. M., Mohning, M. P., **Arbet, J.**, Kroehl, M., Visser, A., August, J., Thomas, S. M., Charry, L. L., Fleischer, C., Feser, M. L., Frazer?Abel, A. A., Norris, J. M., Cherrington, BrianD., Janssen, W. J., Kaplan, M. J., Deane, K. D., Holers, V. M., & Demoruelle, M. K. (2021). Association of sputum neutrophil extracellular trap subsets with IgA anti–citrullinated protein antibodies in subjects at risk for rheumatoid arthritis. *Arthritis & Rheumatology*, *74*(1), 38–48. https://doi.org/10.1002/art.41948

14. Wood, C., **Arbet, J.**, Amura, C. R., Nodine, P., Collins, M. R., Orlando, B. S., Mayer, D. C., Stein, D., & Anderson, J. (2021). Multicenter study evaluating nitrous oxide use for labor analgesia at high- and low-altitude institutions. *Anesthesia & Analgesia*, *134*(2), 294–302. https://doi.org/10.1213/ane.0000000000005712
15. Nodine, P. M., **Arbet, J.**, Jenkins, P. A., Rosenthal, L., Carrington, S., Purcell, S. K., Lee, S., & Hoon, S. (2021). Graduate nursing student stressors during the COVID-19 pandemic. *Journal of Professional Nursing*, *37*(4), 721–728. https://doi.org/10.1016/j.profnurs.2021.04.008
16. **Arbet, J.**, Zhuang, Y., Litkowski, E., Saba, L., & Kechris, K. (2021). Comparing statistical tests for differential network analysis of gene modules. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.630215
17. Ostendorf, D. M., Blankenship, J. M., Grau, L., **Arbet, J.**, Mitchell, N. S., Creasy, S. A., Caldwell, A. E., Melanson, E. L., Phelan, S., Bessesen, D. H., & Catenacci, V. A. (2021). Predictors of long-term weight loss trajectories during a behavioral weight loss intervention: An exploratory analysis. *Obesity Science & Practice*, *7*(5), 569–582. https://doi.org/10.1002/osp4.530
18. Reed, S. M., **Arbet, J.**, & Staubli, L. (2021). Clinical nurse specialists in the united states registered with a national provider identifier. *Clinical Nurse Specialist*, *35*(3), 119–128. https://doi.org/10.1097/nur.0000000000000592
19. Rosenthal, L., Lee, S., Jenkins, P., **Arbet, J.**, Carrington, S., Hoon, S., Purcell, S. K., & Nodine, P. (2021). A survey of mental health in graduate nursing students during the COVID-19 pandemic. *Nurse Educator*, *46*(4), 215–220. https://doi.org/10.1097/nne.0000000000001013
20. Carpenter, C. M., Frank, D. N., Williamson, K., **Arbet, J.**, Wagner, B. D., Kechris, K., & Kroehl, M. E. (2021). tidyMicro: A pipeline for microbiome data analysis and visualization using the tidyverse in r. *BMC Bioinformatics*, *22*(1). https://doi.org/10.1186/s12859-021-03967-2
21. Ramakrishnan, V. R., **Arbet, J.**, Mace, J. C., Suresh, K., Shintani Smith, S., Soler, Z. M., & Smith, T. L. (2021). Predicting olfactory loss in chronic rhinosinusitis using machine learning. *Chemical Senses*, *46*. https://doi.org/10.1093/chemse/bjab042
22. Thomas, E. A., Zaman, A., Cornier, M.-A., Catenacci, V. A., Tussey, E. J., Grau, L., **Arbet, J.**, Broussard, J. L., & Rynders, C. A. (2020). Later meal and sleep timing predicts higher percent body fat. *Nutrients*, *13*(1), 73. https://doi.org/10.3390/nu13010073
23. Schmanski, A., Roberts, E., Coors, M., Wicks, S. J., **Arbet, J.**, Weber, R., Crooks, K., Barnes, K. C., & Taylor, M. R. G. (2020). Research participant understanding and engagement in an institutional, self-consent biobank model. *Journal of Genetic Counseling*, *30*(1), 257–267. https://doi.org/10.1002/jgc4.1316
24. **Arbet, J.**, Brokamp, C., Meinzen-Derr, J., Trinkley, K. E., & Spratt, H. M. (2020). Lessons and tips for designing a machine learning study using EHR data. *Journal of Clinical and Translational Science*, *5*(1). https://doi.org/10.1017/cts.2020.513
25. **Arbet, J.**, McGue, M., & Basu, S. (2020). A robust and unified framework for estimating heritability in twin studies using generalized estimating equations. *Statistics in Medicine*, *39*(27), 3897–3913. https://doi.org/10.1002/sim.8564
26. Coleman-Minahan, K., Sheeder, J., **Arbet, J.**, & McLemore, M. R. (2020). Interest in medication and aspiration abortion training among colorado nurse practitioners, nurse midwives, and physician assistants. *Women's Health Issues*, *30*(3), 167–175. https://doi.org/10.1016/j.whi.2020.02.001
27. Gance-Cleveland, B., Linton, A., **Arbet, J.**, Stiller, D., & Sylvain, G. (2020). Predictors of overweight and obesity in childhood cancer survivors. *Journal of Pediatric Oncology Nursing*, *37*(3), 154–162. https://doi.org/10.1177/1043454219897102
28. James-Allan, L. B., **Arbet, J.**, Teal, S. B., Powell, T. L., & Jansson, T. (2019). Insulin stimulates GLUT4 trafficking to the syncytiotrophoblast basal plasma membrane in the human placenta. *The Journal of Clinical Endocrinology & Metabolism*, *104*(9), 4225–4238. https://doi.org/10.1210/jc.2018-02778
29. Grinde, K. E., **Arbet, J.**, Green, A., O'Connell, M., Valcarcel, A., Westra, J., & Tintle, N. (2017). Illustrating, quantifying, and correcting for bias in post-hoc analysis of gene-based rare variant tests of association. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00117
30. **Arbet, J.**, McGue, M., Chatterjee, S., & Basu, S. (2017). Resampling-based tests for lasso in genome-wide association studies. *BMC Genetics*, *18*(1). https://doi.org/10.1186/s12863-017-0533-3
31. Greco, B., Hainline, A., **Arbet, J.**, Grinde, K., Benitez, A., & Tintle, N. (2015). 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*, *24*(5), 767–773. 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

Batch effects: problem and solution

Boutros Lab, Cancer Data Sciences, University of California, Los Angeles, CA *2025*

- `https://jarbet.github.io/presentations/2025-05-08_batch_effects_nowhite.pdf`

p-value histograms

Boutros Lab, Cancer Data Sciences, University of California, Los Angeles, CA *2025*

- `https://jarbet.github.io/presentation-pvalue-histograms/`

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

The Landscape of Prostate Tumour Methylation *Poster*

UCLA Health Jonsson Comprehensive Cancer Center Annual Symposium, Los Angeles, CA *2025*

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

R

- Over **14 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://uclahs-cds.github.io/package-PrCaMethy/`
  - `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`
- 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