MIN WOO SUN

minwoos@stanford.edu | minwoosun.github.io

EDUCATION

Stanford University *September 2020 - Present*
Biomedical Informatics PhD.
Advisor: Robert Tibshirani

University of California, Los Angeles (UCLA) *September 2013 - December 2017*
Statistics B.S. and Economics B.A. (GPA: 3.69)
Coursework: *Computation and Optimization for Statistics, Data Analysis and Regression, Design and Analysis of Experiment, Linear Algebra, Linear Models, Mathematical Statistics, Monte Carlo Methods, Probability, Statistical Models and Data Mining, Economic Forecasting, Forecasting Exchange Rate, Applied Geo-statistics*

RESEARCH EXPERIENCE

Department of Biomedical Data Science, Stanford University *September 2020 - Present*
Graduate Researcher, advised by Robert Tibshirani

· Develop nested cross-validation algorithm for computing valid confidence intervals for cross-validation estimate of the Cox model test error such as the C-index.
· Develop statistical framework for processing and evaluating a novel MspJI endonuclease based genome-wide DNA methylation quantification in collaboration with the West Lab.
· Collect international Covid-19 data and deploy statistical models to uncover variables that can explain the vast difference in excess mortality rate between the United States and other countries.

Wall Lab, Stanford University *June 2017 - September 2020*
Bioinformatics Research Intern, advised by Dennis Wall and Stefano Moretti

· Developed game theoretic algorithms as a variant-disease association method and implement coalitional game theory on whole genome sequencing data to identify novel candidate genes for autism
· Applied unsupervised machine learning methods to 1965 whole genomes from multiplex families to look for clusters of genotypes differentiating individuals diagnosed with autism from neurotypical individuals
· Run statistical methods such as inter-rater reliability and intraclass correlation to measure the agreement and consistency between different video scorers

Luskin School of Public Affairs, UCLA *October 2016 - May 2017*
Research Assistant, advised by Darin Christensen and Graeme Blaire

· Conducted raster and buffer analysis on geospatial data using R and Amazon EC2 instance
· Implemented GDAL dependency using command line tools to rasterize polygon data faster than existing rgdal package (from 5 hours to less than an hour)
· Integrated Google Maps API to geocode data on crowd instances in South Africa used for buffer analysis

WORK EXPERIENCE

Invitae *January 2018 - August 2020*
Bioinformatics Scientist, Production Bioinformatics

· Implemented statistical methods to optimize lab workflows using Picard and Illumina sequencing metrics
· Computed the likelihood for the various failures modes in the lab to fine-tune the retry logic, reducing the turnaround time for mainline assays
· Built and deployed a Python tool for investigating unusual sex chromosome aneuploidy events attributed to misalignment and rare nondisjunction events
· Automated manual QC reviews for Next Generation Sequencing (NGS) assays and phased out superfluous checks in the bioinformatics pipeline to scale with growing volume
· Evaluated variants to minimize the risk of spurious calls

PREPRINTS & WORKS UNDER REVIEW

* first / co-first author

1. * **Sun, M. W.**, & Tibshirani, R. (2022). Confidence intervals for the Cox model test error from cross-validation. https://doi.org/10.48550/arXiv.2201.10770. [SUBMITTED].

2. Varma, M., Washington, P., Chrisman, B., Kline, A., Leblanc, E., Paskov, K., Stockham, N., Jung, J.-Y., **Sun, M. W.**, & Wall, D. P. (2021). Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile application (p. 2021.06.20.21259187). medRxiv.

3. Washington, P., Kalantarian, H., Kent, J., Husic, A., Kline, A., Leblanc, E., Hou, C., Mutlu, C., Dunlap, K., Penev, Y., Varma, M., Stockham, N., Chrisman, B., Paskov, K., **Sun, M. W.**, Jung, J.-Y., Voss, C., Haber, N., & Wall, D. P. (2020). Training an Emotion Detection Classifier using Frames from a Mobile Therapeutic Game for Children with Developmental Disorders. https://doi.org/10.48550/arXiv.2012.08678

PUBLICATIONS

* first / co-first author

1. Washington, P., Kalantarian, H., Kent, J., Husic, A., Kline, A., Leblanc, E., Hou, C., Mutlu, O. C., Dunlap, K., Penev, Y., Varma, M., Stockham, N. T., Chrisman, B., Paskov, K., **Sun, M. W.**, Jung, J.-Y., Voss, C., Haber, N., & Wall, D. P. (2022). Improved Digital Therapy for Developmental Pediatrics Using Domain-Specific Artificial Intelligence: Machine Learning Study. JMIR Pediatrics and Parenting, 5(2), e26760.

2. Varma, M., Paskov, K. M., Chrisman, B. S., **Sun, M. W.**, Jung, J.-Y., Stockham, N. T., Washington, P. Y., & Wall, D. P. (2021). A maximum flow-based network approach for identification of stable noncoding biomarkers associated with the multigenic neurological condition, autism. BioData Mining, 14(1), 28.

3. Paskov, K., Jung, J.-Y., Chrisman, B., Stockham, N. T., Washington, P., Varma, M., **Sun, M. W.**, & Wall, D. P. (2021). Estimating sequencing error rates using families. BioData Mining, 14(1), 27.

4. Chrisman, B. S., Paskov, K., Stockham, Nate., Tabatabaei, K., Jung, J.-Y., Washington, P., Varma, M., **Sun, M. W.**, Maleki, S., & Wall, D. P. (2021). Indels in SARS-CoV-2 occur at template-switching hotspots. BioData Mining, 14(1), 20.

5. * **Sun, M. W.**, Moretti, S., Paskov, K. M., Stockham, N. T., Varma, M., Chrisman, B. S., Washington, P. Y., Jung, J.-Y., & Wall, D. P. (2020). Game theoretic centrality: A novel approach to prioritize disease candidate genes by combining biological networks with the Shapley value. BMC Bioinformatics, 21(1), 356.

6. Washington, P., Leblanc, E., Dunlap, K., Penev, Y., Kline, A., Paskov, K., **Sun, M. W.**, Chrisman, B., Stockham, N., Varma, M., Voss, C., Haber, N., & Wall, D. P. (2020). Precision Telemedicine through Crowdsourced Machine Learning: Testing Variability of Crowd Workers for Video-Based Autism Feature Recognition. Journal of Personalized Medicine, 10(3), 86.

7. Washington, P., Leblanc, E., Dunlap, K., Penev, Y., Varma, M., Jung, J.-Y., Chrisman, B., **Sun, M. W.**, Stockham, N., Paskov, K. M., Kalantarian, H., Voss, C., Haber, N., & Wall, D. P. (2020). Selection of trustworthy crowd workers for telemedical diagnosis of pediatric autism spectrum disorder. In Biocomputing 2021 (pp. 1425). WORLD SCIENTIFIC.

8. * **Sun, M. W.**, Gupta, A., Varma, M., Paskov, K. M., Jung, J.-Y., Stockham, N. T., & Wall, D. P. (2019). Coalitional Game Theory Facilitates Identification of Non-Coding Variants Associated With Autism. Biomedical Informatics Insights, 11, 1178222619832859.

9. * Gupta, A., **Sun, M. W.**, Paskov, K. M., Stockham, N. T., Jung, J.-Y., & Wall, D. P. (2017). Coalitional game theory as a promising approach to identify candidate autism genes. In Biocomputing 2018 (pp. 436447). WORLD SCIENTIFIC.

PRESENTATIONS

Stanford DBDS Fest 2022. "Confidence intervals for the Cox model test error from cross-validation". Poster.

Molecular Psychiatry 2019. Game Theoretic Centrality Supports Link Between Autism Spectrum Disorder and the HLA Complex. Poster.

The American Society of Human Genetics (ASHG) 2019, Bioinformatics and Computational Approaches. Game Theoretic Centrality Supports Link Between Autism Spectrum Disorder and the HLA Complex. Poster.

American Medical Informatics Association (AMIA) 2018, Translational Bioinformatics. Coalitional game theory as a promising approach to identify candidate autism genes. Poster.

Pacific Symposium on Biocomputing (PSB) 2018, Precision Medicine: From Diplotypes to Disparities Towards Improved Health and Therapies. Coalitional game theory as a promising approach to identify candidate autism genes. Oral presentation.

TEACHING

Stanford Data Science for Social Good
Technical Mentor   *Jun 2022 - Present*

· I am mentoring 3 statistics and data science graduate student fellows on a research project. The program aims to train future researchers to tackle data science projects with social impact. I also created and lectured on commandline / BASH for the Data Science fellows.

BIOMEDIN 217: Translational Bioinformatics
Teaching Assistant   *Jan 2022 - Mar 2022*

· I was a graduate teaching assistant for the course BIOMEDIN 217: Translational Bioinformatics taught by Professor Dennis Wall. I held office hours, graded coursework, and facilitated course discussion for over 80 undergraduate and graduate students. This course is part of the core curriculum in the Biomedical Informatics program at Stanford University.

LEADERSHIP EXPERIENCE

UCLA Residential Life
Resident Assistant   *August 2016 - December 2017*

· Implemented educational and diversity programs for 85 residents
· Trained in peer counseling, conflict mediation, disaster/emergency response, and suicide prevention (QPR Gatekeeper)

Aleph Undergraduate Research Journal
Editor in Chief   *June 2015 - June 2016*

· Oversaw the decisions and work of all 25 staff members for annual publication
· Maintained routine contact with writers on the progress of research articles in various stages: submission, review, revision, and preparation for publication

UCLA UniCamp
Counselor   *March 2015 - June 2015*

· Selected as a camp counselor for 6 high school campers from under-served communities partnered with Project Grad
· Planned and implemented 7 interactive programs to help children pursue higher education in college
· Raised $480 to support camp programming and scholarship for children from low-income families

TECHNICAL SKILLS

Programming languages and Softwares	**Python** (pytorch, numpy, scipy, pandas, scikit-learn, matplotlib), **R** (dplyr, caret, data.table, bioconductor, igraph, ggplot2), **bash**, **SQL**, **bedtools**, **igv**, **QGIS**, **git**, **LaTeX**
Linguistic	English (Native), Korean (Native), Spanish (DELE B1 Certification), Arabic (Elementary)