Alan Aw, PhD

Postdoctoral Researcher
Department of Genetics
University of Pennsylvania

🌍 `https://alanaw1.github.io`
✉ `alan.aw(at)pennmedicine.upenn.edu`
CV last updated: March 1, 2026

EDUCATION

University of California, Berkeley 2018 - 2023
PhD *with Designated Emphasis in Computational and Genomic Biology*, Department of Statistics

- Dissertation title: "Statistical Genomics Through the Lens of Exchangeability, Stability and Stratification"
- GPA: 3.94 (Coursework in theoretical statistics, high-dimensional statistics, applied statistics, convex optimization, statistical genomics, statistical phylogenetics, and social epidemiology)

Stanford University 2014 - 2018
BS *with Honors*, Mathematical and Computational Science

- Honors thesis in Mathematical Evolution and Population Genetics

École normale supérieure de Lyon, France Summer 2014
Modern Mathematics International Summer School for Students

RESEARCH

Note: an asterisk (*) indicates equal contribution.

1. **A.J. Aw** (2026+). Asymptotic enumeration of admixed arrays and a different independence heuristic. *Submitted.*
2. J.V. Moreno-Mayar*, **A.J. Aw***, Y. Deng*, P. Grebenyuk*, …, G. Kroonen, D. Meltzer, Y.S. Song, E. Willerslev (2026+). Ethnohistorical genomes reveal dynamic Holocene origins of Indigenous Siberians. *Under revision.*
3. R. Mandla*, Z. Shi*, K. Hou, Y. Wang, G. Mies, **A.J. Aw**, S. Cullina, Penn Medicine Biobank, E. Kenny, I. Mathieson, E. Atkinson, A.R. Martin, B. Pasaniuc (2026+). Large-scale admixture mapping in the All of Us Research Program improves the characterization of cross-population phenotypic differences. *Under revision.* [PMID: 40236441]
4. **A.J. Aw**, J.F. McRae, E. Rahmani, & Y.S. Song (2026+). Highly parameterized polygenic scores may be learning noisy effect sizes that capture population structure. *Under revision.* (Shiny app) [PMID: 38352303]
5. **A.J. Aw**, R. Mandla, Z. Shi, Penn Medicine Biobank, B. Pasaniuc, & I. Mathieson (2025). Hidden structure in polygenic scores and the challenge of disentangling ancestry interactions in admixed populations. *Genetics* **231**(4): iyaf213. [PMID: 41043808]
6. F. Zhou*, **A.J. Aw***, D. Erdmann-Pham*, J.R. Fischer, & Y.S. Song (2025). Robust and adaptive non-parametric tests for detecting general distributional shifts in gene expression. *Cell Reports Methods* **5**(9): 101147. (Github) [PMID: 40902592]
7. G. Benegas, C. Albors*, **A.J. Aw***, C. Ye*, & Y.S. Song (2025). A DNA language model based on multispecies alignment predicts the effects of genome-wide variants. *Nature Biotechnology.* [PMID: 39747647]

8. **A.J. Aw**, J.P. Spence, & Y.S. Song (2024). A simple and flexible test of sample exchangeability with applications to statistical genomics. *Annals of Applied Statistics* **18**(1): 858-881. (Github) [PMID: 38784669]
9. **A.J. Aw**, L.C. Jin, N.M. Ioannidis, & Y.S. Song (2023). The impact of stability considerations on genetic fine-mapping. *eLife Reviewed Preprint #88039.* (Shiny app) [PMID: 37090514]
10. T.C. Zeng*, **A.J. Aw***, & M.W. Feldman (2018). Cultural hitchhiking and competition between patrilineal kin groups explain the post-Neolithic Y-chromosome bottleneck. *Nature Communications* **9**(1): article no. 2077. [PMID: 29802241]
11. **A.J. Aw** & N.A. Rosenberg (2018). Bounding measures of genetic similarity and diversity using majorization. *Journal of Mathematical Biology* **77**(3): 711-737. [PMID: 29569105]
12. **A.J. Aw** & C.Y. Ku (2015). The covering radius problem for sets of 1-factors of the complete uniform hypergraphs. *Discrete Mathematics* **338**(6): 875-884.
13. **A.J. Aw** (2014). The multicovering radius problem for some types of discrete structures. *Designs, Codes and Cryptography* **72**(2): 195-209.
14. **A.J. Aw** (2012). The Turán number and probabilistic combinatorics. *The American Mathematical Monthly* **119**(6): 510-513.

WORK EXPERIENCE

Bioinformatics Intern 8/2022-6/2023
Illumina, Inc.

- Investigated population structure-related biases in UK Biobank polygenic scores
- Preprint available on bioRXiv: `https://www.biorxiv.org/content/10.1101/2024.01.27.577589v1`
- Manuscript under revision

Statistical Geneticist Intern 5/2021-7/2021
23andMe, Inc.

- Implemented algorithms leveraging properties of multivariate Gaussian distributions to cut down GWAS summary statistic imputation time from 12 hours to 30 minutes
- Incorporated hyperparameter tuning to optimize imputation, and diagnosed performance trade-offs between choices of hyperparameters
- Applied algorithms to downstream genome-wide association studies involving millions of 23andMe customers (e.g., identifying regional hits for phenotypes)

Undergraduate Research Assistant 6/2015-8/2018
Departments of Biology and Statistics, Stanford University

Editorial and Marketing Intern 3/2014-7/2014
Asian Scientist Magazine (now part of Wildtype Media Group)

Research Intern 1/2010-2/2012
Various Research Institutions in Singapore

LEADERSHIP

Workshop Facilitator (Penn Biomedical Graduate Studies) 2023, 2024, 2025
Lead topical discussions for the Experimental Design Workshop for first year Genomic and Computational Biology PhD students, held in August every year.

Reading Group Organizer (Nilah Ioannidis Lab) 1/2021-5/2021
Online reading group focused on fine-mapping. Facilitate discussion and prepare paper reading lists.

Managing Editor (Stanford Undergraduate Research Journal) 2016-2017
Oversaw a team of student writers, editors and marketers to communicate faculty and student research in creative ways. Redesigned website and implemented creative ways of communicating research.

Co-President (Singaporeans at Stanford) 2015-2016
Planned and executed activities to help Singaporean students stay connected within Stanford and with the larger Bay Area Singaporean network. Lead Organizer for Chinese New Year dinner involving 100 guests. Student representative at Prime Minister's visit to the Bay Area as well as the 2016 Singapore-US Trade and Technology Engagement Roundtable.

SERVICE TO COMMUNITY

Penn Medicine Biobank (PMBB) 9/2025-2/2026
I led a team of trainees under the labs of Bogdan Pasaniuc and Iain Mathieson to generate local ancestry calls across about 10^6 imputed and genotyped variants in about 10^4 admixed individuals, for use by PMBB researchers. We prepared a detailed report showcasing applications of our calls while also ensuring all QC steps are reproducible.

Postdoctoral Community 6/2024-6/2025
I helped organize lunches for postdocs in the Department of Genetics. I was part of the Penn Postdoctoral Association Fundraising Committee, where I volunteered for the Science and Technology Fair and mined insights from data collected at previous fairs.

Research Mentorship 8/2021-8/2023
As a graduate student in the Song Lab, I mentored several students (Xurui Rachel Chen, Fanding Zhou) to build open-source software for flexible non-parametric two-sample tests for single-cell RNA-seq data.

Academic Mentorship 5/2020-5/2023
I mentored three undergraduate Statistics majors by providing advice on coursework and career opportunities. Under my mentorship and on top of his already superb capabilities, one student, Wenhao Pan, successfully navigated coursework in computer science and mathematics, as well as undergraduate research opportunities with the Departments of IEOR, Sociology, and Statistics. He is currently a PhD student in Statistics at the University of Washington.

Peer-Reviewing
Reviewer for: *Cell Genomics*, *Health Services Research*, *Nature Communications*, *Statistical Applications in Genetics and Molecular Biology*, and *Mathematical Reviews*

SKILLS

Programming Languages: R (including interfacing with C++), Python.

Tools: Git, Bash, GNU Multiple Precision Library.

Bioinformatics Software: PLINK, GCTA, SuSiE, Polyfun, LDPred, momi2, msprime, ms, SLiM, Seurat, SKAT and other rare variant association testing methods, RFMix.

Languages: Bilingually native in Chinese, and can understand basic French.

SOFTWARE

R: flintyR, QRscore

Python: flintyPy, zagar

TEACHING

Berkeley:

- Statistical Computing Facility R Bootcamp Instructor (Summer 2022)
    - Answer questions one on one pertaining to various computational statistics topics. Teach module on Calculations (topics include: vectorization, use of `apply` and variants, merging and joining, stratified analyses)
- COMPSCI198: Algorithmic Fairness and the Genome (Fall 2021)
    - Gave a guest lecture about the negative impact of neglected LD pattern heterogeneities between populations on the identification of causal variants and on the construction of equitable polygenic scores
    - Organized guest lecture by 23andMe scientist, Dr. James Ashenhurst, to share efforts by 23andMe to reduce healthcare disparities
- Stat 153: Introduction to Time Series (Spring 2020)
- Stat 135: Concepts of Statistics (Spring 2019)

AWARDS & HONORS

NIH T32 Grant Awardee (Genomic Medicine), Bethesda, MD	2025
Invited Session Speaker, STATGEN 2025, Minneapolis, MN	2025
ASHG Reviewer's Choice Abstract (Top 10% of Posters), Denver, CO	2024
JSM Travel Award, SF Bay Area Chapter of the American Statistical Association, CA	2022
Outstanding GSI Award, Berkeley, CA	2021
Mortimer Fleishhacker Scholarship, Berkeley, CA	2019
Dean's Award for Academic Achievement, Stanford, CA	2018
Best Poster, 23andMe Genome Research Day, Mountain View, CA	2018
Bio-X Undergraduate Research Fellowship, Stanford, CA	2016