

# Alan J. Aw, PhD

✉ alan.aw(at)pennmedicine.upenn.edu • 🌐 alanaw1.github.io

CV last updated: July 9, 2026

## Education

---

### University of California, Berkeley

*PhD, Department of Statistics*

2018–2023

Designated Emphasis in Computational and Genomic Biology

Dissertation: *Statistical Genomics Through the Lens of Exchangeability, Stability and Stratification*

*Selected coursework:* Stat 210AB (Theoretical Statistics), Stat 215AB (Applied Statistics), EECS 227C (Convex Optimization), ELENG 290 (High-Dimensional Data Analysis), Stat 243 (Statistical Computing), Stat C245C/F (Statistical Genomics)

### Stanford University

*BS with Honors, Mathematical and Computational Science*

2014–2018

Honors thesis in Mathematical Evolution and Population Genetics

*Selected coursework:* Stat 310AB (Theory of Probability), Math 236 (Stochastic Differential Equations), MS&E 322 (Stochastic Calculus and Control), Math 159 (Discrete Probabilistic Methods)

### École normale supérieure de Lyon, France

*Modern Mathematics International Summer School for Students*

Summer 2014

## Positions

---

### Department of Genetics, University of Pennsylvania

*Postdoctoral Researcher*

2023–present

Advisors: Iain Mathieson (primary) and Bogdan Pasaniuc

## Research and Publications

---

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**\*, P. Grebenyuk\*, Y. Deng\*, . . . , G. Kroonen, D. Meltzer, Y.S. Song, E. Willerslev (2026+). Ethnohistorical genomics uncover the 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. *Accepted in principle at Nature Communications*. [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. Albers\*, **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.

## Awards & Honors

---

**2026:** Invited Speaker, UC Santa Cruz Applied Math Department, CA  
**2025–2027:** NIH T32 Grant Awardee (Genomic Medicine), Bethesda, MD  
**2025:** Invited Session Speaker, STATGEN 2025, Minneapolis, MN  
**2024:** ASHG Reviewer’s Choice Abstract (Top 10% of Posters), Denver, CO  
**2022:** JSM Travel Award, SF Bay Area Chapter of the American Statistical Association, CA  
**2021:** Outstanding GSI Award, Berkeley, CA  
**2019:** Mortimer Fleishhacker Scholarship, Berkeley, CA  
**2018:** Dean’s Award for Academic Achievement, Stanford, CA  
**2018:** Best Poster, 23andMe Genome Research Day, Mountain View, CA  
**2016:** Bio-X Undergraduate Research Fellowship, Stanford, CA

## Teaching

---

### Statistical Computing Facility, UC Berkeley

*R Bootcamp Instructor*

*Summer 2022*

### COMPSCI198: Algorithmic Fairness and the Genome, UC Berkeley

*Guest Lecturer*

*Fall 2021*

- Lectured on the negative impact of neglected LD pattern heterogeneities between populations on causal variant identification and equitable polygenic score construction
- Organized guest lecture by Dr. James Ashenurst (23andMe) on reducing healthcare disparities

### Stat 153: Introduction to Time Series, UC Berkeley

*Graduate Student Instructor*

*Spring 2020*

### Stat 135: Concepts of Statistics, UC Berkeley

*Graduate Student Instructor*

*Spring 2019*

## Work Experience

---

### illumina, Inc.

*Bioinformatics Intern*

*8/2022–6/2023*

- Investigated population structure-related biases in UK Biobank polygenic scores

### **23andMe, Inc.**

*Statistical Geneticist Intern* 5/2021–7/2021

- Implemented algorithms leveraging multivariate Gaussian properties to reduce GWAS summary statistic imputation time from 12 hours to 30 minutes
- Incorporated hyperparameter tuning to optimize imputation and diagnose performance trade-offs
- Applied algorithms to downstream GWAS involving millions of 23andMe customers

### **Departments of Biology and Statistics, Stanford University**

*Undergraduate Research Assistant* 6/2015–8/2018

### **Asian Scientist Magazine, Singapore**

*Editorial and Marketing Intern* 3/2014–7/2014

### **Various Research Institutions, Singapore**

*Research Intern* 1/2010–2/2012

## **Leadership**

---

### **Penn Biomedical Graduate Studies**

*Workshop Facilitator* 2023, 2024, 2025

Led topical discussions for the Experimental Design Workshop for first-year Genomic and Computational Biology PhD students.

### **Nilah Ioannidis Lab, UC Berkeley**

*Reading Group Organizer* 1/2021–5/2021

Online reading group on fine-mapping: facilitated discussion and prepared paper reading lists.

### **Stanford Undergraduate Research Journal**

*Managing Editor* 2016–2017

Oversaw team of student writers, editors and marketers; redesigned journal website.

### **Singaporeans at Stanford**

*Co-President* 2015–2016

Organized community events including Chinese New Year dinner (100 guests); student representative at Prime Minister's Bay Area visit and the 2016 Singapore-US Trade and Technology Engagement Roundtable.

## **Service to Community**

---

### **Master's Thesis, École normale supérieure de Lyon**

*External Examiner* 4/2026–6/2026

Evaluated literature review and final report of Lorraine Bocquet-Appel.

### **Penn Medicine Biobank (PMBB)**

*Data Resource Contribution* 9/2025–5/2026

Led a team of trainees (Mathieson and Pasaniuc labs) to generate local ancestry calls across  $\sim 10^6$  variants in  $\sim 10^4$  admixed individuals; presented findings at PMBB Monthly Meeting (May 2026).

### **Department of Genetics, University of Pennsylvania**

*Postdoctoral Community* 6/2024–6/2025

Organized postdoc lunches; volunteered for Penn Postdoctoral Association Science and Technology Fair.

### **Song Lab, UC Berkeley**

*Research Mentor* 8/2021–8/2023

Mentored Xurui Rachel Chen and Fanding Zhou on building open-source software for non-parametric two-sample tests for single-cell RNA-seq data.

### **UC Berkeley**

*Academic Mentor* 5/2020–5/2023

Advised three undergraduate Statistics majors on coursework and career opportunities; mentee Wenhao Pan is now a PhD student in Statistics at the University of Washington.

**Peer Review:** Reviewer for *Annals of Applied Statistics*, *Cell Genomics*, *Health Services Research*, *Nature Communications*, *Statistical Applications in Genetics and Molecular Biology*, *Mathematical Reviews*

## Skills

---

**Programming:** R (including C++ interfacing), Python

**Tools:** Git, Bash, GNU Multiple Precision Library

**Bioinformatics:** Verily Workbench (Cloud Computing), PLINK, GCTA, SuSiE, Polyfun, LDpred, momi2, msprime, ms, SLiM, Seurat, SKAT, RfMix

**Languages:** Bilingually native in Chinese; basic French

## Software

---

**R:** flintyR, QRscore

**Python:** flintyPy, zaggar

## Professional Affiliations

---

- American Mathematical Society
- American Society of Human Genetics
- American Statistical Association
- Genetics Society of America
- Society for Modeling and Theory in Population Biology