# Yusuph Mavura, PhD, MS

San Francisco, CA • 470 399 1865 • yrmavura@gmail.com

## EDUCATION

- **PhD, Epidemiology and Translational Science**, University of California, San Francisco (2019-2025)
- **MS, Bioinformatics,** Georgia Institute of Technology (2017-2018)
- **BS, Biochemistry & Molecular Biology,** Jomo Kenyatta University of Agriculture & Technology (2012-2016)

## SKILLS

- **Computational Tools:** Python, R, Bash script, Machine learning, MacOS, Windows, Unix
- **Statistical Analysis:** Linear, logistic, Cox, longitudinal mixed effects models; non-parametric tests; Bayesian statistics; dose–response; imputation; machine learning; power calculations
- **Genomic Data Analysis:** exome/genome sequencing analysis; population structure (PCA), ancestry, relatedness, consanguinity analyses; GWAS; polygenic scores; rare disease genetics.
- **Real-World Data Analysis:** Extract, clean and harmonize longitudinal electronic health records (clinical, lab, prescription, genetic, survey data); high performance computing pipelines for big data processing
- **Leadership & Mentorship:** NIH F99/K00 grant writing and project management; interdisciplinary coordination; GitHub code management; mentor junior researchers

## SUMMARY AND QUALIFICATIONS

- Motivated genetic epidemiologist, bioinformatician trained in advanced statistical and causal inference methods aiming to apply skills to large-scale real-world data in industry
- Technical expertise in genome sequencing data and statistical analysis of complex longitudinal electronic health records data, resulting in multiple peer-reviewed publications and 1 NIH-funded grant
- Experienced in large-scale nationwide, cross-disciplinary research initiatives, coordinating geneticists, clinicians, statisticians, and database managers across multiple institutions.
- Strong written and verbal science communication skills, with multiple manuscripts, including *NEJM*

## RESEARCH AND PROFESSIONAL EXPERIENCE

**Graduate Student Researcher;** University of California, San Francisco                    2019-2025
*Project 1: Dissertation work on assessing clinical utility of multi-ancestry polygenic scores*

- Leveraged longitudinal real-world electronic health records (EHR) data to assess the clinical utility of multi-ancestry polygenic scores
- Designed and implemented pipelines integrating Kaiser Permanente genotype and EHR data using R/Python/Bash, and high-performance computing (HPC)
- Led and designed analyses of environmental, and genetic, effects on antihypertensive monotherapy response in a Kaiser Permanente dataset with ~100k individuals
- Authored three dissertation manuscripts and maintained all analysis code in GitHub for reproducibility, version control and collaboration

*Project 2: Rare/Mendelian disease genomics and diagnostic yield*
- Assessed diagnostic yield of exome and genome sequencing, finding no evidence of differences by genetic ancestry in cases with suspected mendelian disorders.
- Processed and cleaned genetic, clinical, survey data from the Clinical Sequencing Evidence-Generating Research (CSER) Phase II consortium for analysis
- Designed and implemented genetic ancestry, relatedness, and consanguinity estimation analyses for multiple nationwide consortia studies

*Project 3: Prostate cancer genomics in men of African ancestry*
- Led a research project characterizing the genomic rearrangement landscape in localized prostate cancer samples from Nigerian men
- First author manuscript coordinating and collaborating with researchers in Nigeria

**Intern;** National Cancer Institute (NCI), National Institutes of Health (NIH)          Jun-Aug 2018
- Conducted a Genome-Wide Association study (GWAS) which identified multiple new risk loci for pancreatic cancer in East Asian populations
- Developed and executed data pipelines in PLINK and R for population-stratification