YUBO (KRISTY) MA

651-363-5737 ◇ kma1@unc.edu ◇ 1100 W NC Hwy 54 BYP, Chapel Hill, NC

EDUCATION

M.S. in Biostatistics. University of North Carolina at Chapel Hill — Expected May 2025
Certificate in Innovation, Leadership and Management.
One of the top programs providing highly rigorous and theoretical training for PhD in Biostatistics.

B.A. in Statistics. Macalester College — Aug 2019 - May 2023
Minors: Computer Science, Biology   GPA: 3.94 (Major: 4.00) - Summa Cum Laude
Kofi Annan International Student Scholarship recipient   Dean's List
Coursework: Intermediate Statistical Methods, SAS programming, Survival Analysis (Planned), Statistical theory (Graduate Level), Causal Inference, Algorithm, Computational Biology, Team Collaboration

EXPERIENCE

Statistical Genetics Researcher (Advisor: Dr. Xihao Li) — Present
Department of Biostatistics, University of North Carolina — *Chapel Hill, NC*

- Developing computational frameworks and add-on customizations to STAAR, an established R package testing variant-set association using annotation information in whole-genome sequencing studies from TOPMed database

Computational Biology Researcher — Sep 2022 - Dec 2022
Department of Plant and Microbial Biology, University of Minnesota — *Minneapolis, MN*
Worked on annotated whole genome sequencing data on transposable elements and gene expression portfolio

- Constructed machine learning framework and automation scripts in R and Shell to perform classification for 10K+ transposable elements with **85%** accuracy
- Dynamically visualized transposable elements annotation data in 26 maize lines using **R shiny** to assess frequency of suspected mutation events
- Evaluated **20%** errors including typos, unstated assumptions, and potential limitations to strengthen the statistical rigor of the study
- Inspected the putative insertion regions and test how TE insertions correlated with differential expression of nearby genes

Translational Genomics Researcher — Jun 2022 - Aug 2022
Department of Agronomy and Plant Genetics, University of Minnesota — *Saint Paul, MN*

- Carried out multiple GWAS analyses versus flowering times, yield, and moisture content, finding out two significant structural variations associated with late flowering time across 100+ maize domestic inbred lines
- Collected and validated data for a support vector machine model predicting moisture contents of nixtamalized maize, then visualized clustering of genotypes using PCA for 300+ hybrid maize genotypes in R

Teaching and Lab Assistant — Sep 2021 - May 2023
Statistic, Biology department, Macalester College — *Saint Paul, MN*

- Assisted 200 students' learning in Machine Learning, Calculus II, Molecular Genetics, and Epidemiology
- Performed series of molecular biology lab techniques through the semesters
- Hold weekly office hours to work personally with students to strengthen their understanding of statistics; Prepared review material, proctored and evaluated exams, lab projects, homework, and quizzes

PROJECTS

Project 1 Statistical Genetics: Applying mixed effect model in correlated genetic data — GitHub

- Conducted a small GWAS study involving simulation of genotypes from real genetic data, testing liner mixed model's effectiveness in correcting family correlation structure
- Self-taught multiple R packages (GENESIS, SNPRelate, snpStats, GWASTools) to clean a handle large HapMap data (**145M** SNPs), reducing noise by **25%** and running cost by **18%**
- Filtered out 120 uncorrelated individuals from the original data set of **2000**-individual pool from correlation coefficient matrix between pairs using KING tool set, resulting 6% increase in the precision of the research outcomes

Project 2 Machine Learning: Breast cancer data — GitHub

- Extracted 600+ observations from Kaggle database and performed data wrangling; Took on **70%** of highly analytical report writing

- Built models and evaluated the performance of GAM, Random Forest, LASSO and K-means predicting the area and malignancy of tumor using 30 variables
- Demonstrated high correlation for 4 predictors with tumor malignancy and implemented diagnostic matrix and ROC AUC graphs evidenced model optimization, achieving overall accuracy of **97%**

Project 3 Analytical Writing: Literature review on genetic variants on SARS-CoV-2   Link

- Integrated and summarized the findings of 20+ high-quality published papers discussing various effects genetic variants of SARS-CoV-2
- Identified 13 mutant strains and 30+ specific amino acid or genetic mutations and loci related to variants' prevalence
- Thoroughly discussed the variants' impacts from aspects including respective characteristics, transmissibility, and clinical morbidity, also critically analyzed the limitations of investigation such as higher bias in epidemiology

PUBLICATION AND PRESENTATIONS

- *In progress:* Manisha Munasinghe, **Kristy Ma**, et.al. **Combined analysis of transposable elements and structural variation in maize genomes reveals genome contraction outpaces expansion.**
- Poster presentation: **Translational Genomics Research: Exploring Maize**. Summer 2022 Research Symposium. Macalester College, Saint Paul, MN, Sep 2022
- **1st Place on Demo Business Case Competition.** Financial Service (HK) Association. July 2019.
- **Scientific Communication Capstone presentation Gold Award.** Macalester College. March 2023.

SKILLS

Programming Languages	R, Git, SAS, Python, Java, Unix Shell Scripting
R tools	tidyverse/ggplot2/Bioconductor/R markdown/ R bookdown/ R Shiny
Teaching	Fundamentals of Epidemiology, Machine Learning, Study Design in Public Health
Others	2-year leadership for a dance organization