Arham Salman

281-650-5936 | arham.salman@uth.tmc.edu | https://arham4124.github.io/ | linkedin.com/in/arham-salman12d/

EDUCATION

The University of Texas Health Science Center Houston | GPA: 4.0 May 2026

Master of Science in Biostatistics & Data Science

Coursework: Intermediate Biostatistics, Intro to Data Science, Data Analytics and Predictions, Applied Linear Regression, Categorical Data Analysis, Probability and Distribution Theory

The University of Texas at Austin | GPA: 3.3 May 2024

Bachelor of Science in Public Health

Coursework: Public Health, Calculus I, Calculus II, Linear Algebra, Python Programming, Epidemiology.

EXPERIENCE

MD Anderson Cancer Center, Chong Wu Lab June 2025 – Present

Graduate Research Assistant

- Engineered Evo2 DNA Foundational Model using Tensorflow and Pytorch to focus training on single nucleotide polymorphism (SNP), improving validation loss for variant prediction tasks.
- Achieved 91% accuracy in predicting SNPs in chromosome 22 validation split using next token prediction in Genotype-Tissue Expression data.
- Developed retrieval-augmented generation (RAG) architecture through Kubernetes and LangChain on hugging face Llama 2-7b model to retrieve information from vector embeddings upon user inputs.

UTHealth Houston, Texas SPAN Project June 2025 – Aug 2025

Data Analyst Intern

- Created comprehensive data dictionaries for the Texas School Physical Activity and Nutrition (SPAN) and analyzed obesity trends in 2,000+ students across Texas by cleaning data through SQL and producing logistic regression models using SAS.
- Discovered significance in mother's education level and how safe parents feel about walking their kids to school as predictors of class 1 obesity ($p<0.05$) through logistic regression models and chi-square test.

UTHealth San Antonio, Liang Ma Lab March 2025 – Aug 2025

Graduate Research Assistant

- Generated Python permutation test and matplotlib visualizations that prioritize single nucleotide polymorphism signal heights for gene CYP2D6 in Chromatin Immunoprecipitation Sequencing (ChIP-seq) data related to Alzheimer's Disease (AD).
- Integrated FORGEdb from National Institute of Health (NIH) to further filter SNPs based on chromatin region activity scores.
- Prioritized SNPs rs133381, rs40822088, and rs2143139 as significantly enriched ($p<0.0001$) in chromatin, offering novel insights for AD.
- Utilized Find Individual Occurring Motifs (FIMO) to generate motifs that displayed transcription factors associated with rs2143139.

UTMB Summer Institute of Biostatistics & Data Science Jun 2023 – Jul 2023

Biostatistics Intern

- Conducted team research with principal investigators under NIH funded program and determined crucial biomarkers and their implications on cardiovascular disease using logistic regression and Wilcoxon signed-rank test in R.
- Found statistically significant differences in BP and HBA1C levels among rich and poor household incomes ($p<0.05$).

The University of Texas at Austin, SHIFT Program Jan 2023 - May 2023

Data Analyst Intern

- Created deliverable to determine key stakeholders in Austin to contact about substance use among UT college students.
- Collected over 100 responses through Qualtrics from UT students and built interactive Excel dashboards with pivot charts to visualize substance use patterns, identifying top 3 venues influencing 30% of student alcohol use behavior.

PROJECTS

Risk Factors Associated with Low Birthweight

- Analyzed data from Baystate Medical Center to assess risk factors on low birthweight through descriptive tests, two sample t-tests, and fitted multiple regression model using STATA.
- Results indicated health disparities among people of color and smokers at Baystate Medical Center ($p<0.05$).

PubMed Web Scraper

- Developed web scraper to extract research article details of Cardiovascular research from PubMed using Biopython library.
- Imported CSV file to SQLite to create database where Python was utilized to query publications.
- Created data visualizations using with time series line plot, bar plot, and word cloud to display top journals, frequently used words, and authors in Jupyter Notebook.

PUBLICATIONS

Zhang Z, Zhang W, Salman A, Feng H, Sun R, Zhao B, Lin L, Wu L, Deng H-W, Pan W, Wu C. Practical insights for integrating agentic AI across scientific research workflows.

PRESENTATIONS

Education-Related Behaviors and their impact on Childhood Obesity in Texas GetPHIT SUMMIT 2025
Exploring the Link Between Household Income and Cardiovascular Disease Factors in Adults. UTMB 2023

COMMUNITY INVOLVEMENT

Biostatistics Student Association 2025

Secretary

- Created Kahoot trivia game for introductory social and promoted inclusivity by socializing with members.

Society of Asian Scientists and Engineers (SASE) 2022 - 2024

Family Head

- Organized and led weekly social events for over 40 members, tailored to diverse interests and schedules.
- Created and captained SASE intramural basketball team with over 20 players
- Played for SASE intramural volleyball team and led team to playoffs, averaging 3 points a game.

AWARDS

Runner Up Best Poster Award – GetPHIT Summit 2025
Outstanding Public Health Scholar – University of Texas Health Science Center Houston 2024

SKILLS

Technical /Computer Skills: Python, R, STATA, SAS, SQL, Microsoft Word, Microsoft Excel.