

Ibrahim Hossain Sajal



Rockville, MD



(682)-232-8791



sajali2@nih.gov



ihsajal.github.io

About Me

I am a statistician with a PhD in Data Science & Statistics, with 5+ years of experience in statistical genetics and scalable Bayesian modeling to drive biomarker discovery and cancer risk prediction in biomedical research.

Education

Ph.D. in Data Science & Statistics

University of Texas at Dallas
Richardson, TX, USA (2019 - 2024)

Master's. in Applied Statistics

University of Dhaka
Dhaka, Bangladesh (2017 - 2018)

Bachelor's in Applied Statistics

University of Dhaka
Dhaka, Bangladesh (2013 - 2016)

Skills

Programming & Computing:

R, Python, SAS, SQL, HPC, Linux commands

Statistical Genetics:

GWAS, Curation & Harmonization, Mendelian Randomization, Colocalization, Instrumental Variable, Mediation Analysis, Differential Gene Expression, DNANexus

Statistical Learning:

Survival Analysis, Conditional & Penalized Regression (LASSO), Variable Selection, Cross-Validation, Multiple Imputation, Bayesian Inference, MCMC, Latent Variable Models

Machine Learning:

Supervised & Unsupervised Learning, Decision Trees, Bagging, Random Forest, Boosting, SVM, KNN, Variational Inference

Awards

- Fellows Award for Research Excellence
- Mei Lein Fellowship
- Boyd Harshburger Travel Award
- ABGOD Conference Travel Award
- PhD Research Small Grant Program
- SISG Scholarship
- Academic Excellence Award

Experience

Postdoctoral Fellow

National Cancer Institute, NIH (2025 - Present)

- Identified therapeutic targets linking risk factors to kidney cancer by employing Mendelian Randomization and colocalization analyses on proteomics summary statistics from UK Biobank (UKB-PPP).
- Developed a scalable, summary-statistics-based multivariate testing framework to improve detection of genetic associations.
- Analyzed pan-cancer phenotypes and protein networks to uncover pathway-level signals relevant to target discovery (KEGG, Hallmark).
- Conducted survival analyses in lung cancer to evaluate prognostic impact of clinical risk factors and polygenic risk scores.

Graduate Research Assistant

The University of Texas at Dallas (2020 - 2024)

- Designed haplotype-based genetic association test, **bivariate Quantitative Bayesian LASSO**, for correlated disease phenotypes.
- Extending Bayesian LASSO models to high-dimensional, settings using variational inference to enable scalable genetic discovery.
- Developed **CBCRisk-Black** and **CBCRisk-Mastectomy**, two personalized absolute risk prediction tools to support clinical decision-making in contralateral breast cancer.
- Analyzed population level registry datasets (BCSC, SEER) to identify risk factors and estimate cancer incidence and hazard rates.

Publications

- **Sajal, I.H.**, Song, A.J., Brown, K.M., Machiela, M.J., Kraft, P., Chanock, S., Dutta, D., (2026). Prioritizing plasma proteomic mediators of the effects of risk factors on renal cell carcinoma. (Under Review).
- **Sajal, I. H.**, & Biswas, S. (2023). Bivariate quantitative Bayesian LASSO for detecting association of rare haplotypes with two correlated continuous phenotypes. *Frontiers in Genetics*, 14, 1104727.
- **Sajal, I. H.**, Chowdhury, M., Wang, T., Euhus, D., Choudhary, P. K., & Biswas, S. (2022). CBCRisk-Black: a personalized contralateral breast cancer risk prediction model for black women. *Breast Cancer Research and Treatment*, 194(1), 179-186.

Presentations

- Joint Statistical Meetings (2025) — Invited Talk
- NIH Research Festival (2025) — Poster
- Advances in Statistical and Computational Methods for Analysis of Biomedical, Genetic, and Omics Data (2023) — Poster
- Southern Regional Council on Statistics (2022) — Poster

Work Authorization: Approved EAD, Permanent Residency Pending