

Overcoming biases in molecular causal inference

Sajal Kumar¹, and Mingzhou Song^{1,2}

¹Department of Computer Science, ²Molecular Biology and Interdisciplinary Life Sciences Graduate Program, New Mexico State University

Contact email: sajal49@nmsu.edu, joemsong@nmsu.edu

Computer inference of biological mechanisms is increasingly approachable due to dynamically rich data sources such as single-cell genomics. Inferred molecular interactions can prioritize hypotheses for wet-lab experiments to expedite biological discovery. However, complex data often come with unwanted biological or technical variations, exposing biases over marginal distribution and sample size in current methods to favor spurious causal relationships. Here, we present an adapted functional chi-squared test (AdpFunChisq) that rewards functional patterns over non-functional or independent patterns. On synthetic and three biology datasets, we demonstrate the advantages of AdpFunChisq over ten methods on overcoming biases that give rise to wide fluctuations in the performance of alternative approaches. On single-cell multiomics data of multiple phenotype acute leukemia, we found that T-cell surface glycoprotein CD3 delta chain (CD3D) may causally mediate specific genes in the viral carcinogenesis pathway. AdpFunChisq offers a viable option for robust causal inference in complex dynamical systems.