

Our paper on [Gene network inference by fusing data from diverse distributions](#) has been published in Bioinformatics. We will present it at [ISMB 2015 in Dublin](#).

In the paper we describe FuseNet, a Markov network formulation that infers networks from a collection of potentially nonidentically distributed datasets.

Markov networks are undirected graphical models that are widely used to infer relations between genes from experimental data. Their state-of-the-art inference procedures assume the data arise from a Gaussian distribution. High-throughput omics data, such as that from next generation sequencing, often violates this assumption. Furthermore, when collected data arise from multiple related but otherwise nonidentical distributions their underlying networks are likely to have common features. New principled statistical approaches are needed that can deal with different data distributions and jointly consider collections of datasets.

FuseNet is computationally efficient and general: given any number of distributions from an exponential family, FuseNet represents model parameters through shared latent factors that define neighborhoods of network nodes. In a simulation study, we demonstrate good predictive performance of FuseNet in comparison to several popular graphical models. We show its effectiveness in an application to breast cancer RNA-sequencing and somatic mutation data, a novel application of graphical models. Fusion of datasets offers substantial gains relative to inference of separate networks for each dataset. We also demonstrate that network inference methods for non-Gaussian data can help in accurate modeling of the data generated by emergent high-throughput technologies.