

Generative Model in Words:

1. Assume a set of latent variables representing isoform compositions across multiple samples.
2. Specify a probabilistic model that captures the generation of short-read RNA-seq data, accounting for the challenges of isoform identification, such as shared subsequences and sample-specific frequencies.
3. Incorporate a Bayesian nonparametric approach that does not rely on known isoform reference sequences but estimates isoform composition directly using a shared catalog across samples.
4. Develop a stochastic variational inference method for efficient and robust posterior inference.

Generative Model in Math (Simplified and Hypothetical):

Prior:

Isoform Composition \sim Bayesian Nonparametric Prior (latent variables)

Likelihood:

RNA-seq Data \sim Probabilistic Model for Short-Read Data

(captures challenges like shared subsequences and sample-specific frequencies)