Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > stat > arXiv:1106.3211

Search or Article-id

(Help | Advanced search)

All papers | Go!

Statistics > Methodology

Statistical Modeling of RNA-Seq Data

Julia Salzman, Hui Jiang, Wing Hung Wong

(Submitted on 16 Jun 2011)

Recently, ultra high-throughput sequencing of RNA (RNA-Seq) has been developed as an approach for analysis of gene expression. By obtaining tens or even hundreds of millions of reads of transcribed sequences, an RNA-Seq experiment can offer a comprehensive survey of the population of genes (transcripts) in any sample of interest. This paper introduces a statistical model for estimating isoform abundance from RNA-Seq data and is flexible enough to accommodate both single end and paired end RNA-Seq data and sampling bias along the length of the transcript. Based on the derivation of minimal sufficient statistics for the model, a computationally feasible implementation of the maximum likelihood estimator of the model is provided. Further, it is shown that using paired end RNA-Seq provides more accurate isoform abundance estimates than single end sequencing at fixed sequencing depth. Simulation studies are also given.

Submission history

From: Julia Salzman [view email]

[v1] Thu, 16 Jun 2011 12:17:05 GMT (330kb)

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.