P30

TIGAR2: sensitive and accurate estimation of transcript isoform expression with longer RNA-Seq reads

Naoki Nariai, Kaname Kojima, Takahiro Mimori, Yukuto Sato, Yosuke Kawai, Yumi Yamaguchi-Kabata, Masao Nagasaki

Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Japan

Abstract

Background

High-throughput RNA sequencing (RNA-Seq) enables quantification and identification of transcripts at single-base resolution. Recently, longer sequence reads become available thanks to the development of new types of sequencing technologies as well as improvements in chemical reagents for the Next Generation Sequencers. Although several computational methods have been proposed for quantifying gene expression levels from RNA-Seq data, they are not sufficiently optimized for longer reads (e.g. > 250 bp).

Results

We propose TIGAR2, a statistical method for quantifying transcript isoforms from fixed and variable length RNA-Seq data. Our method models substitution, deletion, and insertion errors of sequencers based on gapped-alignments of reads to the reference cDNA sequences so that sensitive read-aligners such as Bowtie2 and BWA-MEM are effectively incorporated in our pipeline. Also, a heuristic algorithm is implemented in variational Bayesian inference for faster computation. We apply TIGAR2 to both simulation data and real data of human samples and evaluate performance of transcript quantification with TIGAR2 in comparison to existing methods.

Conclusions

TIGAR2 is a sensitive and accurate tool for quantifying transcript isoform abundances from RNA-Seq data. Our method performs better than existing methods for the fixed-length reads (100 bp, 250 bp, 500 bp, and 1000 bp of both single-end and paired-end) and variable-length reads, especially for reads longer than 250 bp.