Pacific Center for Emerging Infectious Diseases Research

UNIVERSITY of HAWAI'I MĀNOA

Department of Tropical Medicine, Medical Microbiology & Pharmacology

JOHN A BURNS SCHOOL OF MEDICINE, UNIVERSITY OF HAWAI'I AT MANOA

Uncovering the Splicing Signatures for ENCODE RNA-Seq Data using Read-Split-Fly

Two typical molecular machineries: the major spliceosome (U2-dependent) and the minor spliceosome (U12-type) are involved in pre-mRNA splicing. It is generally thought that most canonical or non-canonical splicing events involving U2- and U12 spliceosomes occur within nuclear pre-mRNAs. However, the question of whether at least some U12-type splicing occurs in the cytoplasm is still unsettled. What is clear is that at least one special type of splicing occurs in the cytoplasm: *Ire1α* dependent splicing of *Xbp1* mRNA. This form of splicing is inducible under conditions of endoplasmic reticulum (ER) stress.

In recent years next-generation sequencing technologies have revolutionized the field. The "Read-Split-Walk" (RSW) and "Read-Split-Run" (RSR) methods were developed to identify genome-wide non-canonical spliced regions including special events occurring in cytoplasm. As a significant amount of genome/transcriptome data such as, Encyclopedia of DNA Elements (ENCODE) project, have been generated, we have advanced a newer more memory-efficient version of the algorithm, "Read-Split-Fly" (RSF), which can detect non-canonical spliced regions with higher sensitivity and improved speed. The RSF algorithm also outputs the spliced sequences for further downstream biological function analysis.

We used open access ENCODE project RNA-Seq data to search spliced intron sequences against the well-known orthologous U12-type spliceosomal intron database U12DB. Preliminary results of searching 70 ENCODE samples indicated that the presence of 5' splice sites with U12-type signature is more frequent than U2-type and prevalent in noncanonical junctions reported by RSF. Specifically, we reported that both U12-type and U2-type intron 5' splice sites queries hit more novel splicing junctions relative to the known splice junctions, whereas both U12-type and U2-type 3' splice sites queries hit less novel splicing junctions relative to the known splice junctions. We also observed that there are more U12-type than U2-type for the 5' splice sites category. The selected spliced sequences have also been further studied using miRBase to elucidate their functionality. Several miRNAs are prevalent in studied ENCODE samples. Two of these (*hsa-miR-1273* and *hsa-miR-548*) are associated with many diseases as suggested in the literature.

Our RSF pipeline is able to detect many possible junctions (especially those with a high expression) with very high overall accuracy and relative high accuracy for novel junctions. We suggest RSF, a tool for identifying novel splicing events, is applicable to study a range of diseases across biological systems under different experimental conditions.

Yongsheng Bai, Ph.D.

Assistant Professor of Bioinformatics
Department of Biology
The Center for Genomic Advocacy (TCGA)
Indiana State University
Terre Haute, Indiana

Monday, January 9, 2017 at 11:00 a.m.
John A. Burns School of Medicine, Kaka'ako Campus
Medical Education Building Auditorium (Room 315)
For further information, contact (808) 692-1654

This seminar is supported by grant P30GM114737 (COBRE), P20GM103466 (INBRE), U54MD007584 (RMATRIX), and G12MD007601 (BRIDGES) from the National Institutes of Health.