

AS-EAST: a functional annotation tool for putative proteins encoded by alternatively spliced transcripts

Masafumi Shionyu^{1,*}, Ken-ichi Takahashi¹ and Mitiko Go^{1,2}

¹Department of Computer Bioscience, Faculty of Bioscience, Nagahama Institute of Bio-Science and Technology, 1266, Tamura-cho, Nagahama, Shiga 526-0829 and ²Research Organization of Information and Systems, 4-3-13 Toranomon, Minato-ku, Tokyo 105-0001, Japan

Associate Editor: Burkhard Rost

ABSTRACT

Summary: Alternative Splicing Effects Assessment Tools (AS-EAST) is an online tool for the functional annotation of putative proteins encoded by transcripts generated by alternative splicing (AS). When provided with a transcript sequence, AS-EAST identifies regions altered by AS events in the putative protein sequence encoded by the transcript. Users can evaluate the predicted function of the putative protein by inspecting whether functional domains are included in the altered regions. Moreover, users can infer the loss of inter-molecular interactions in the protein network according to whether the AS events affect interaction residues observed in the 3D structure of the reference isoform. The information obtained from AS-EAST will help to design experimental analyses for the functional significance of novel splice isoforms.

Availability: The online tool is freely available at <http://as-alps.nagahama-i-bio.ac.jp/ASEAST/>.

Contact: m_shionyu@nagahama-i-bio.ac.jp

Received on March 8, 2012; revised on April 30, 2012; accepted on May 25, 2012

1 INTRODUCTION

In higher eukaryotes, genes often produce alternatively spliced transcripts (AS transcripts). Many AS events have been actively detected with high-throughput experimental methods, such as RNA-Seq and microarrays (Hallegger *et al.*, 2010). However, the functions of putative proteins encoded by AS transcripts (termed ‘AS isoforms’) have not been experimentally analyzed in many cases. Functional annotation tools for transcripts with novel patterns of splicing are desirable to infer the functional significance of AS isoforms. There are few tools for analyzing AS isoforms translated from novel transcripts queried by users. AltAnalyze (<http://www.altanalyze.org>) identifies AS events using RNA-Seq or microarray data and shows how these events may affect domain composition. However, it does not provide information on the effects of AS on the 3D structures of AS isoforms. MAISTAS (Floris *et al.*, 2011) assesses whether user-queried AS isoforms are structurally plausible proteins, but explicit functional annotations are not provided.

Previously, we developed a pipeline that detects regions altered by AS events (termed ‘AS regions’) in AS isoforms using genome sequences and full-length transcript data (Yura *et al.*, 2006). The pipeline then evaluates the impact of AS events on the

interactions between the AS isoforms and other molecules by identifying interaction residues from 3D structure data of relevant molecular complexes. All of the data derived from the pipeline are provided in the AS-ALPS database (Shionyu *et al.*, 2009). In this article, we describe AS-EAST that annotates and analyzes user-uploaded transcript sequences using AS-ALPS. AS-EAST determines whether the transcript encodes a novel AS isoform and annotates such functional sites in the AS isoform as residues interacting with other molecules. We provide an example: AS-EAST predicts that a novel AS isoform of mitogen-activated protein kinase 1 (MAPK1) in human skeletal muscle inhibits the signaling pathway by removing residues that interact with ATP and substrate proteins.

2 OVERVIEW OF AS-EAST

2.1 Input data

To detect and annotate AS events in a user-submitted transcript (termed ‘query transcript’), AS-EAST accepts a FASTA-formatted transcript sequence. A novel pattern of splicing detected with RNA-Seq or exon junction microarray often determines whether a certain exon in a known transcript model is skipped. AS-EAST has a user interface for generating an exon-skipped transcript sequence from the known transcript sequences stored in AS-ALPS. For example, RNA-Seq data (Wang *et al.*, 2008) shows that the fourth exon of the MAPK1 transcript tends to be skipped in human skeletal muscle (Fig. 1a). This splicing pattern is novel because no fourth exon-skipped transcript is found in RefSeq or Ensembl transcript datasets stored in AS-ALPS. Users can build a fourth exon-skipped transcript sequence by checking the checkbox of ‘exon 4’ and selecting the ‘Generate’ button. Then, the transcript sequence excluding the fourth exon is shown in FASTA format.

2.2 AS region detection

First, a genome contig sequence aligned to the query transcript sequence with the largest value of both length coverage and sequence identity is selected using MEGABLAST (Altschul *et al.*, 1997). An alignment of the query transcript and the contig sequence is performed using SPLIGN (Kapustin *et al.*, 2008). AS-EAST searches the AS-ALPS database for transcripts mapped on the same region of the contig as the query transcript. The user can select one of the transcripts as a reference transcript. Second, the protein-coding sequence (CDS) of the query transcript is predicted by identifying the longest open reading frame (ORF) or FrameDP program (Gouzy *et al.*, 2009). Users can also use the ORF starting

*To whom correspondence should be addressed.

