Genomics Proteomics Bioinformatics 12 (2014) 105–109

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

APPLICATION NOTE

PrimerSeq: Design and Visualization of RT-PCR Primers for Alternative Splicing Using RNA-seq Data

Collin Tokheim [#], Juw Won Park, Yi Xing [*]

Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA

Received 19 March 2014; revised 4 April 2014; accepted 8 April 2014
Available online 18 April 2014

Handled by Jun Yu

KEYWORDS

Alternative splicing;
RNA-seq;
Primer design;
Transcriptome;
Visualization

Abstract The vast majority of multi-exon genes in higher eukaryotes are alternatively spliced and changes in alternative splicing (AS) can impact gene function or cause disease. High-throughput RNA sequencing (RNA-seq) has become a powerful technology for transcriptome-wide analysis of AS, but RT-PCR still remains the gold-standard approach for quantifying and validating exon splicing levels. We have developed PrimerSeq, a user-friendly software for systematic design and visualization of RT-PCR primers using RNA-seq data. PrimerSeq incorporates user-provided transcriptome profiles (*i.e.*, RNA-seq data) in the design process, and is particularly useful for large-scale quantitative analysis of AS events discovered from RNA-seq experiments. PrimerSeq features a graphical user interface (GUI) that displays the RNA-seq data juxtaposed with the expected RT-PCR results. To enable primer design and visualization on user-provided RNA-seq data and transcript annotations, we have developed PrimerSeq as a stand-alone software that runs on local computers. PrimerSeq is freely available for Windows and Mac OS X along with source code at http://primerseq.sourceforge.net/. With the growing popularity of RNA-seq for transcriptome studies, we expect PrimerSeq to help bridge the gap between high-throughput RNA-seq discovery of AS events and molecular analysis of candidate events by RT-PCR.

Introduction

Alternative splicing (AS) is a prevalent mechanism of gene regulation in higher eukaryotes [1]. AS plays an important role in both normal biological processes [2] and disease [3]. In recent years, high-throughput RNA sequencing (RNA-seq) has become a powerful and popular technology for global analysis of AS [4]. From the massive amount of RNA-seq reads, one can discover novel splicing events, quantify the usage level of

[*] Corresponding author.
E-mail: yxing@ucla.edu (Xing Y).
[#] Present address: Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

Production and hosting by Elsevier

1672-0229/$ - see front matter © 2014 Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China. Production and hosting by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gpb.2014.04.001

alternatively spliced exons in any RNA sample of interest, and identify genome-wide changes in AS between different biological states. However, RT-PCR is still regarded as the most reliable and standard approach to quantify and validate exon splicing levels [5]. In fact, researchers customarily perform RT-PCR validation of AS events discovered from RNA-seq data prior to downstream functional studies.

A widely used measure of AS is the percent-spliced-in (PSI, or ψ) metric, which measures the percent inclusion level of an alternatively spliced exon (or splice site) in the final mRNA products [4]. In an RNA-seq project, the PSI value of an AS event can be first estimated from RNA-seq data using software like mixture-of-isoforms (MISO) [4] or multivariate analysis of transcript splicing (MATS) [6] and then validated independently by RT-PCR. However, the design of RT-PCR primers for AS analysis is typically a time-consuming step that requires tedious manual operations. Software that allow input of a template sequence for primer design like Primer3 [7] and associated interfaces Primer3Web, Primer3Plus and BatchPrimer3 [8] can theoretically design primers for any AS event of interest. However, it is left to the user to manually extract sequences for primer design, which is time-consuming and error-prone. This is particularly challenging for RNA-seq projects, where researchers may need to validate tens or even hundreds of AS events identified from the transcriptome-wide AS analysis (for example, see [9,10]). Consequently validation of AS from big RNA-seq data has become a major bottleneck between high-throughput discovery of AS events and in-depth analysis of molecular function and regulation.

It should also be noted that the repertoire of expressed genes and mRNA isoforms is dynamically regulated, while current tools and databases for RT-PCR primer design use static (pre-defined) gene and transcript annotations and do not incorporate transcriptome profiles for the specific sample(s) of interest. Primer databases like GETPrime [11], RTPrimerDB [12], PrimerBank [13] and qPrimerDepot [14] only contain primers for a restricted set of species and are built on pre-defined gene annotations. Most primer design tools (*e.g.*, PerlPrimer [15] and QuantPrime [16]) or primer databases (mentioned above) focus on gene expression and occasionally transcript expression rather than AS events. In general, there is a lack of primer design tools or databases for AS analysis, with a few exceptions being GETPrime (gene and transcript specific) [11] and RASE (alternative splicing) [17]. RASE is the method most specifically designed for AS analysis, but its web interface only works with human genes and requires time-consuming manual input of sequences. It should also be noted that RNA-seq is a flexible technology, which can be applied to any organism of interest. In fact, researchers have used RNA-seq to study AS in a wide variety of organisms such as honey bee [18], silkworm [19], *Plasmodium falciparum* [20] and frog [21]. Additionally, computational tools such as Cufflinks [22] and Scripture [23] can be used to construct transcript annotations *de novo* from RNA-seq data aligned to the genome. Therefore, an ideal primer design tool for AS analysis should have the flexibility to work with user-provided RNA-seq data on a diverse range of organisms, instead of being restricted to a small set of species and pre-defined transcript annotations.

Here we present PrimerSeq, a user-friendly stand-alone software for systematic design and visualization of RT-PCR primers for AS analysis. PrimerSeq has a graphical user interface (GUI) and "one-click" type installation for convenient access by a wide range of users. It utilizes user-provided RNA-seq data to define splicing patterns, estimate exon inclusion levels (PSI, or ψ), select suitable regions for placement of RT-PCR primers and visualize RNA-seq data along with highlighting expected RT-PCR results. Users can conveniently compare the graphical output of PrimerSeq to their RT-PCR experimental result.

Methods

PrimerSeq workflow and algorithm

PrimerSeq designs RT-PCR primers for AS analysis. The design process can incorporate the transcriptome profiles of the samples of interest through user-provided RNA-seq data files, or only utilize pre-defined gene and transcript annotations. As shown in the flow diagram (**Figure 1**), the input to PrimerSeq includes a genome sequence file (FASTA), a gene and transcript annotation file (GTF), mapped RNA-seq reads (BAM, recommended but optional) and a list of exon coordinates representing the events of interest. Visualizing read density also requires a BigWig file, although this visualization step is optional. For each AS event, PrimerSeq attempts to place a pair of forward and reverse PCR primers on suitable flanking exons. Such flanking exons can be specified by users in the input. Alternatively, PrimerSeq can automatically choose

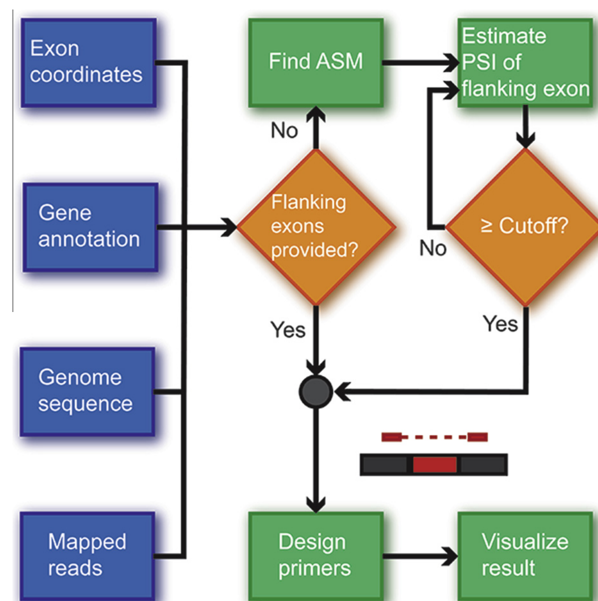

Figure 1 The flow diagram of PrimerSeq
PrimerSeq flow diagram designates inputs as blue, computations as green and decisions as orange. If flanking exons are specified by the user, PrimerSeq will immediately design primers. If not specified, PrimerSeq will first identify the alternative splicing module (ASM) containing the target exon and then iteratively search for the closest flanking exons above a user-defined PSI (ψ) cutoff. Results are subsequently visualized through plotting the RNA-seq data juxtaposed with the expected RT-PCR results, which include estimated ψ values for the target exon. Visualizing read density, an optional feature, requires a BigWig file. PSI stands for percent-spliced-in.

Download English Version:

https://daneshyari.com/en/article/2822465

Download Persian Version:

https://daneshyari.com/article/2822465

Daneshyari.com