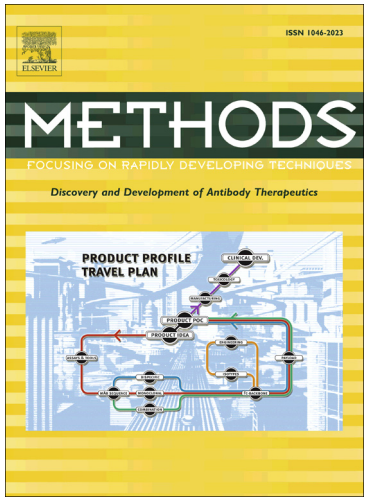
Accepted Manuscript



Pushing the Annotation of Cellular Activities to a Higher Resolution: Predicting Functions at the Isoform Level

Wenyuan Li, Chun-Chi Liu, Shuli Kang, Jian-Rong Li, Yu-Ting Tseng, Xianghong Jasmine Zhou

PII: S1046-2023(15)30037-2
DOI: http://dx.doi.org/10.1016/j.ymeth.2015.07.016
Reference: YMETH 3756

To appear in: *Methods*

Received Date: 1 June 2015
Revised Date: 20 July 2015
Accepted Date: 29 July 2015

Please cite this article as: W. Li, C-C. Liu, S. Kang, J-R. Li, Y-T. Tseng, X.J. Zhou, Pushing the Annotation of Cellular Activities to a Higher Resolution: Predicting Functions at the Isoform Level, *Methods* (2015), doi: http://dx.doi.org/10.1016/j.ymeth.2015.07.016

Pushing the Annotation of Cellular Activities to a Higher Resolution: Predicting Functions at the Isoform Level

Wenyuan Li [1 *], Chun-Chi Liu [2 3 4 *], Shuli Kang [1 *], Jian-Rong Li [2] and Yu-Ting Tseng [2], and Xianghong Jasmine Zhou [1 ξ]

[1] Molecular and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA
[2] Institute of Genomics and Bioinformatics, National Chung Hsing University, Taiwan
[3] Institute of Biomedical Sciences, National Chung Hsing University, Taiwan
[4] Agricultural Biotechnology Center, National Chung Hsing University, Taiwan

* Joint first authors
ξ To whom correspondence should be addressed. Tel: +1 213 740 7055; Fax: +1 213 740 2475; Email: xjzhou@usc.edu

Abstract

In past decades, the experimental determination of protein functions was expensive and time-consuming, so numerous computational methods were developed to speed up and guide the process. However, most of these methods predict protein functions at the gene level and do not consider the fact that protein isoforms (translated from alternatively spliced transcripts), not genes, are the actual function carriers. Now, high-throughput RNA-seq technology is providing unprecedented opportunities to unravel protein functions at the isoform level. In this article, we review recent progress in the high-resolution functional annotations of protein isoforms, focusing on two methods developed by the authors. Both methods can integrate multiple RNA-seq datasets for comprehensively characterizing functions of protein isoforms.

1. Introduction

Alternative splicing of pre-mRNAs is a key mechanism for increasing transcriptome and proteome complexity in eukaryotic cells. More than 90% of human multi-exon genes undergo alternative splicing [1,2]. The proteins translated from alternatively spliced mRNA have different amino acid sequences and structures, and can have distinct (even opposing) biological functions. Alternative splicing plays important roles in development, physiology, and a large number of human diseases including cancer [3–5]. It is also reported to be a driver of the evolution of phenotypic complexity in mammals [1,6,7]. Because of the limitations of current experimental techniques, there are very few functional annotations of protein isoforms available in the