TAPAS: tools to assist the targeted protein quantification of human alternative splice variants

Jae-Seong Yang1,2, Eduard Sabido2,3, Luis Serrano1,2,4,* and Christina Kiel1,2,*

1EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), 2Universitat Pompeu Fabra (UPF), 3Proteomics Unit, Centre for Genomic Regulation (CRG), 08003 Barcelona and 4Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

ABSTRACT

Motivation: In proteomes of higher eukaryotes, many alternative splice variants can only be detected by their shared peptides. This makes it highly challenging to use peptide-centric mass spectrometry to distinguish and to quantify protein isoforms resulting from alternative splicing events.

Results: We have developed two complementary algorithms based on linear mathematical models to efficiently compute a minimal set of shared and unique peptides needed to quantify a set of isoforms and splice variants. Further, we developed a statistical method to estimate the splice variant abundances based on stable isotope labeled peptide quantities. The algorithms and databases are integrated in a web-based tool, and we have experimentally tested the limits of our quantification method using spiked proteins and cell extracts.

Availability and implementation: The TAPAS server is available at URL http://davinci.crg.es/tapas/.

Contact: luis.serrano@crg.eu or christina.kiel@crg.eu

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on March 21, 2014; revised on June 22, 2014; accepted on June 29, 2014

1 INTRODUCTION

Higher eukaryotic proteomes have greatly expanded during evolution by the variation created by alternatively spliced forms of the same gene, duplication into orthologous genes and differentially processed proteins (Nilsen and Graveley, 2010; Rappsilber and Mann, 2002). These multiple protein isoforms play significant roles in differentiation, development and disease (Irimia and Blencowe, 2002; Maniatis and Tasic, 2002). High-throughput genomic, transcriptomic and mass spectrometry (MS) technologies have facilitated the genome-wide identification of alternative splicing (Nagaraj et al., 2011; Pan et al., 2008). It was shown that shifts in the relative splice transcript abundance between different cells and tissues are more frequently observed than all-or-none switch-like behaviors (Shen et al., 2012). Thus, knowing the amount of an expressed splice variant is highly desirable to understand cellular functions.

Peptide-centric high-throughput shotgun MS approaches have achieved remarkable detection coverage (Nagaraj et al., 2011). However, inferring protein isoform identities and estimating their abundances by peptide-centric shotgun MS is not straightforward, as many peptides are shared between isoforms (Nesvizhskii and Aebersold, 2005). Recently, statistical methods have addressed these problems by using both unique and shared peptides (Blein-Nicolas et al., 2012; Dost et al., 2012; Gerster et al., 2014). While these methods have improved protein quantifications from peptide intensities, they were not specifically designed for a targeted MS approach of isoform analysis. In this work, we developed efficient computational optimization techniques to infer a minimal list of unique and shared peptides needed to accurately quantify a set of isoforms. This algorithm is integrated in a web-based tool (TAPAS) that assists in the design of stable isotope labeled (SIL) peptides for targeted MS experiments.

2 METHODS AND IMPLEMENTATION

TAPAS has been implemented using Django, a web framework based on the Python programming language. With this software, the user will be able to retrieve the optimal experimental design and estimates for absolute protein abundances for a selected set of splicing variants of interest motivated from a specific biological problem (Fig. 1).

Experimental design: TAPAS generates lists of unique and shared peptides that can be used in SIL-based targeted proteomics approaches to quantify selected alternative splice variants. The peptides are specific for one or more input genes, but are not shared with other genes to avoid non-specificity. TAPAS uses two complementary algorithms, recursive set subtraction and Gaussian elimination to generate a minimal combination of peptides required for quantifying each splice variant or a group of splice variants (Supplementary Material). Furthermore, TAPAS provides additional information to the user such that suitable peptides can be chosen: TAPAS warns if a peptide has a possible posttranslational modification or an incomplete peptidase digestion, which could lead to an underestimate of the peptide quantity. Furthermore, it prioritizes peptides based on previous experimental evidence obtained by MS. To perform its function, TAPAS requires inputs for the query genes, the query databases (Swiss-Prot/TrEMBL) and/or the user-defined splice variant sequences.

Absolute quantification of splice variants: TAPAS estimates the amounts of alternative splice variants by considering the abundance of peptides measured from SIL-based targeted MS...
TAPAS is a framework that brings together experimental design and computational analysis for the absolute quantification of splice variants. It designs experiments by selecting the minimal combination of shared and unique peptides needed to analyze a specific set of splice variants. By taking advantage of quantitative SIL-based proteomics, TAPAS can estimate the quantities of splice variants. Our method also allows users to query protein family members with homologous sequences. We tested TAPAS with SRM-based MS approaches, but we anticipate that this splice variant quantification strategy can be applied to other targeted MS approaches, such as multiplexed MS/MS (MSX) and Sequential Window Acquisition of all THEoretical Mass Spectra (SWATH), which would enable numerous splice variants to be quantified in a single experiment.

ACKNOWLEDGEMENTS

Protein expression and purification was done in the CRG Biomolecular Screening & Protein Technologies Unit, and MS-based quantifications, in the CRG/UPF Proteomics Unit. The authors thank Kiana Toufighi and Martin Schaefer for comments on the manuscript. The authors acknowledge support of the Spanish Ministry of Economy and Competitiveness, ‘Centro de Excelencia Severo Ochoa 2013-2017’ (SEV-2012-0208).

Funding: The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement Nr. PRIMES_278568. This work was supported by the Spanish Ministerio de Economía y Competitividad, Plan Nacional BIO2012-39754 and the European Fund for Regional Development. The CRG/UPF Proteomics Unit is part of the “Plataforma de Recursos Biomoleculares y Bioinformáticos” (Instituto de Salud Carlos III), supported by grant PT13/0001.

Conflicts of interest: none declared.

REFERENCES


