Coding Potential of the Products of Alternative Splicing in Human

LEONI G (1, 2), LE PERA L (1), FERRÈ F (1), RAIMONDO D (1), TRAMONTANO A (1, 3)

(1) Dipartimento di Scienze Biochimiche - Sapienza Università di Roma, Rome, Italy
(2) INRAN, Via Ardeatina, 546 - 00178 Roma, Italy
(3) Istituto Pasteur Fondazione Cenci Bolognetti - Sapienza Università di Roma, Rome, Italy

Motivation
Alternative splicing (AS) is a mechanism used by cells to diversify the proteins produced by a gene. Estimates of the amount of AS in human have risen dramatically over recent years, especially since the advent of novel high-throughput sequencing technologies1, 2, 3, reaching up to the 95% of the multi-exon genes 4. While the role of AS in expanding the functional complexity of a genome is established, less clear is whether all generated transcripts do indeed encode for functional proteins and therefore expand the coding potential of a genome.

Methods
Aim of this study is to analyze alternative splicing isoforms of human gene products that are unambiguously identified by mass spectrometry and compare their properties with those of isoforms of the same genes for which no evidence of existence was found in publicly available mass spectrometry datasets. The selected isoforms were analyzed for the presence of uninterrupted functional domains, active sites as well as the plausibility of their predicted structure. We report how well each of these strategies and their combination can correctly identify translated isoforms and derive a lower limit for their specificity, that is, their ability to correctly identify non-translated products.

Results
The most effective strategy for correctly identifying translated products relies on the conservation of active sites, but it can only be applied to a small fraction of isoforms, while a reasonably high coverage, sensitivity and specificity can be achieved by analyzing the presence of non-truncated functional domains. Combining the latter with an assessment of the plausibility of the modeled structure of the isoform increases both coverage and specificity with a moderate cost in terms of sensitivity.

Contact email
guido.leoni@gmail.com