Comparative analysis of the alternative splicing diversity in the human and mouse brain proteomes: preliminary results

Esdras Matheus da Silva[1], Thais Martins[1], Raphael Tavares da Silva[2], Fabio Passetti[3],

1 Oswaldo Cruz Institute
2 Universidade Federal de Minas Gerais
3 FIOCRUZ - IOC

Abstract

The alternative splicing of pre-mRNAs in eukaryotes can generate an extensive complexity of alternative protein variants from a given gene. Some mutations in the genome can cause malfunction of pre-mRNA splicing mechanism and, hence, generate protein variants with the potential to lead to neurodegenerative diseases. Model organisms, particularly, the mouse, are often used for the study of many pathologies because of ethical and methodological convenience. The mouse brain proteome has been investigated to improve the knowledge of the molecular aspects of neurodegenerative diseases. Currently, mass spectrometry (MS) is the most used technology for protein complex sample analysis. This methodology is based on the digestion of a given protein sample followed by their mass detection, as well as their identification by computational analysis with a traditional protein sequence repositories. However, some peptides are not identified by these traditional repositories because they comprise protein variants derived from alternative splicing events. As a solution, proteogenomic approaches have been used to identify these peptides that cannot be found in convencional repositories. In short, this strategy consists in using genome or transcriptome data to build customized protein repositories. Here, we investigated the diversity of protein variants formed by alternative splicing in the human and mouse brain, by using MS public data and two customized protein sequence repositories, one for each species, created by our group. First, we detected alternative splicing variants in complete reference mRNA (RefSeq) data along with ESTs through a methodology developed by our group called ternary matrices. Second, after an in silico translation, the predicted proteins were in silico digested and the resulting peptides were selected if they were not comprise in any sequence from Uniprot/SwissProt database. Third, customized repositories were created based on the union of selected peptides and Uniprot/SwissProt protein sequences. Forth, we made proteomic analysis using these repositories and comparative analysis of results from both species in order to identify the expression of orthologous genes at the protein level. The customized repository for humans had 20,150 canonical sequences and 204,294 nonredundant peptides from protein variants formed by alternative splicing. The mouse repository had 16,888 canonical sequences and 156,889 nonredundant peptides from protein variants formed by alternative splicing. In the MS experiments of the brain's corpus callosum for both species we identified 1,040 orthologous genes expressing canonical proteins and 5 orthologous genes expressing protein variants formed by alternative splicing. We believe that this study can contribute to the better understanding of the alternative splicing profile that can be found in both human and mouse's brains. Financial support: CAPES, FAPERJ and Fiocruz.