

Identification of alternative splicing variants that are susceptible to NMD pathway by a bioinformatic approach

Vinicius da Silva Coutinho Parreira, Letícia Graziela Costa Santos de Mattos,
Fabio Passetti

Laboratory of Gene Expression Regulation, Carlos Chagas Institute, Fundação Oswaldo Cruz (Fiocruz), Curitiba, PR, Brazil

Abstract

DNA high-throughput sequencing associated with bioinformatics approach, permits to perform in silico analysis of genomes and transcriptomes. The last GENCODE release provide 19, 975 genes related to more than 83, 000 proteins. This difference can be associated with RNA post-transcriptional modifications, mainly associated to alternative splicing. Alternative splicing increases transcriptomic diversity due to alternative recognition of intron splice sites. Although alternative splicing is strictly regulated, some errors may occur, such as the insertion of a premature termination codon (PTC) in the transcript and such molecules could result in a truncated protein. When a termination codon is located at least 55 nucleotides downstream of an exon-exon junction, it is called PTC and the transcript is molecularly marked for degradation by the nonsense mediated decay pathway (NMD). The computational prediction of NMD events or the manual annotation is used by the Ensembl project to provide organism-based NMD datasets. SpliceProt is a protein sequence repository created by our research group that aims to predict splice variants and perform their in silico translation. We aim to create a subgroup of SpliceProt that will have a PTC identification for transcripts, consequently removing these variants of in silico translation step. We have updated the SpliceProt repository to receive Ensembl transcript information for human. The rat and mouse datasets is ongoing. Using a ternary matrices methodology, which received mapping coordinated produced by the BLAT software, we have identified 222, 462 human splice variants that corresponds to 61, 122 human genes. According to preliminary results using human known control mRNAs with NMD-targets and non-NMD-targets, we were able to correctly identify all NMD events in concordance to the NMD Ensembl annotation (e.g. LENE and CRYZ). Currently, we are performing the PTC identification in the SpliceProt datasets using the NMDClassifier software. In addition, we will integrate the NMD classifier in the repository's pipeline to enable automate future SpliceProt releases.

Funding: National Council for Scientific and Technological Development (CNPq), Coordination for the Improvement of Higher Education Personnel - Capes, Carlos Chagas Institute - Fundação Oswaldo Cruz (Fiocruz)