Computational Assessment of the Role of Repeat Sequences in Splicing of the Human RNA Transcript During HIV-infection

Karki, Sabin

TDP-43, a protein functioning in transcription and pre-mRNA splicing regulation, is known to cause exon skipping on the CFTR and apoA-II genes by binding to TG-repeat motifs 12 bases or longer at intronic 3' splice sites during transcription. TDP-43 also represses viral transcription of HIV-1, and expression of the protein's associated gene, TARDBP, is part of the host cellular response to HIV-1 infection in humans. This research aimed to computationally identify potential binding sites of TDP-43 on a virally-infected human genome, predicting an increase in pre-mRNA splicing events via exon skipping. A Python program was created to search for all TG-motifs 12 bases or longer within intronic sequences of the human genome, then modified to reduce results so only those TG-sequences within an established limit for the distance from the TG-sequence terminus to the next intron/exon junction would be outputted. Results indicated that the 3' splice site of intron 1 of gene XAF1 represented a potential binding site for TDP-43, as it displayed the presence of a 26-base long TG-motif within the established distance limit. Knowledge of the molecular mechanisms of transcription-related regulatory events of HIV-1 is critical in progress toward a therapeutic cure. Furthermore, research on TDP-43 may implicate it as a potential causal factor in diseases associated with inactivated protein products if the associated gene possesses TG-repeat motifs.