Ovelapping alternative donor splicing sites in the human genome

Ekaterina O. Ermakova, Ramil N. Nurtdinov, and Mikhail S. Gelfand

Alternative donor sites constitute up to 15% of all alternative splicing events in the human genome. In 11% of alternative donor site pairs, the sites overlap and thus challenge the consensus requirements. We consider 2869 pairs of human alternatively spliced overlapping donor sites and 127558 singlet sites that overlap a silent site-like motif. Pairs with the site shift of 4 nucleotides are the most abundant despite the frameshift they introduce. Site usage in pairs is usually uneven, the downstream site is preferred in more than 2/3 of cases. In 61% of pairs, only one of the sites involved in splicing leads to a possibly translated isoform while the other one produces an mRNA isoform which is likely subject to nonsense-mediated decay (NMD). The abundance of overlapping donor sites in human genes is likely to be underestimated. Overlapping alternative donor sites and acceptor sites may have different functional roles: alternative splicing of overlapping acceptor sites mainly leads to microvariations in protein sequences, whereas alternative donor sites often lead to frameshifts and thus either yield major differences in protein sequence and structure, or generate NMD-inducing mRNA isoforms.