Title: Finding Transposon-Sensitive Zones in Mammalian Introns

Author(s): Ying Zhang; Mark Romanish; Dixie L. Mager
Affiliation(s): Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada
Contact email: yzhang@bccrc.ca

Abstract:

Comprising nearly half of the human and mouse genomes, transposable elements (TEs) are found within most genes. Although the vast majority of TEs in introns are fixed in the species and presumably exert no significant effects on the enclosing gene, some markedly perturb transcription and result in disease or a mutated phenotype. Factors determining the likelihood that an intronic TE will affect transcription are not clear. In this study, we examined intronic TE distributions in both human and mouse and found several factors that likely contribute to whether a particular TE can influence gene transcription. Specifically, we observed that TEs near exons are greatly underrepresented compared to random distributions, but the size of these "underrepresentation zones" differs between TE classes. Compared to elsewhere in introns, TEs within these zones are shorter on average and show stronger orientation biases. Moreover, TEs in extremely close proximity (< 20 bp) to exons show a strong bias to be near splice-donor sites. Interestingly, disease-causing intronic TE insertions show the opposite distributional trends, and by examining EST databases, we found that the proportion of TEs contributing to chimeric TE-gene transcripts is significantly higher within their underrepresentation zones. Moreover, computational predictions on the number and strength of potential splice sites in the TE sequences that we have tested showed a significant decrease near intron-exon boundaries. Based on these factors, we selectively examined a list of polymorphic mouse endogenous retroviral elements (ERVs) in introns and showed clear evidence of transcriptional disruption by ERV insertions in the *Trpc6* and *Kcnh6* genes. Taken together, these studies lend insight into the potential selective forces that have shaped intronic TE distributions and enable identification of TEs most likely to exert transcriptional effects on genes.