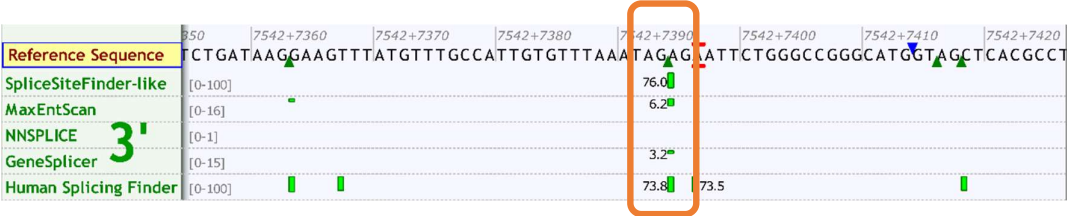
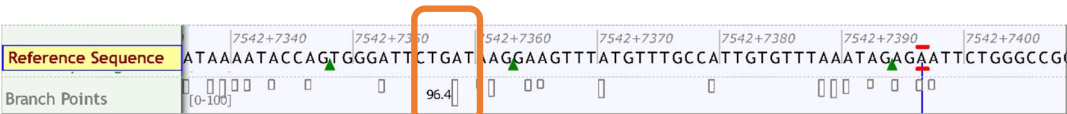


Figure S1. Bioinformatic splice-site and branch-point analysis. To identify candidate regions for the LINE-1 insertion, different algorithms were used to evaluate potential acceptor splice sites and branch points within intron 51, namely the SpliceSiteFinder-Like, MaxEntScan (Yeo et al., 2004), NNSPLICE (Reese et al., 1997), GeneSplicer (Pertea et al., 2001) and Human Splice Finder (Desmet et al., 2009), integrated in Alamut Visual software (v2.8, Interactive Biosoftware, Rouen, France).

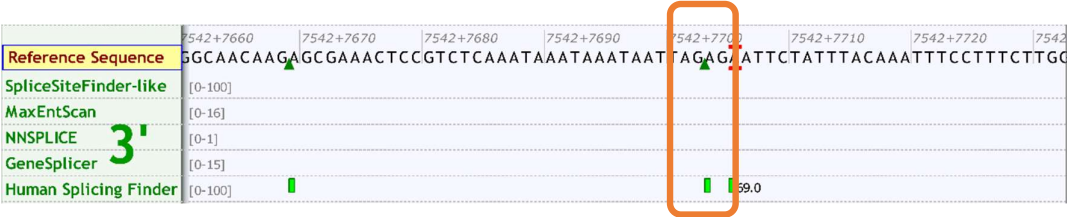
REGION 1 – Potential acceptor splice-site



REGION 1 – Potential branch-points



REGION 2 – Potential acceptor splice-site



REGION 2 – Potential branch-points



REGION 3 – Potential acceptor splice-site

	7542+8910	7542+8920	7542+8930	7542+8940	7542+8950	7542+8960	7542+8970	7542+8980
Reference Sequence	AACAGGATTTGATTGTGCTTTTATGATTTTAAAGATTTCATTAAAAATAATGCCACGGTTTCTAAAATGAT							
SpliceSiteFinder-like	[0-100]							
MaxEntScan	[0-16]							
NNSPLICE	[0-1]							
GeneSplicer	[0-15]							
Human Splicing Finder	[0-100]							

REGION 3 – Potential branch-points

	7542+8870	7542+8880	7542+8890	7542+8900	7542+8910	7542+8920	7542+8930	7542+8940
Reference Sequence	GTAGTAATGGCTGCATTGTTTTGAAAAAAAAAAGCAACAGGATTGATTGTGCTTTTATGATTTTAAAG							
Branch Points	[0-100]							

REGION 4 – Potential acceptor splice-site

	7543-8220	7543-8210	7543-8200	7543-8190	7543-8180	7543-8170	7543-8160	7543-8150
Reference Sequence	TGCTACCTAACTTATTTGCCATTATTTGTAAAGATTCCAGGCTCTAACCAGCCATTGAGGCTTTATT							
SpliceSiteFinder-like	[0-100]							
MaxEntScan	[0-16]							
NNSPLICE	[0-1]							
GeneSplicer	[0-15]							
Human Splicing Finder	[0-100]							

REGION 4 – Potential branch-points

	7543-8260	7543-8250	7543-8240	7543-8230	7543-8220	7543-8210	7543-8200
Reference Sequence	CTCAAACTGATTTTAAAAAGCCAATATATACTGCTTCTACTATGCTACCTAACTTATTTGCCATTATT						
Branch Points	[0-100]						

REGION 5 – Potential acceptor splice-site

	7543-8070	7543-8060	7543-8050	7543-8040	7543-8030	7543-8020	7543-8010	7543-8000
Reference Sequence	TGTAGTGCCATACTTGATATGAAGTCTTTGCTGAGATTCCCATATGTGGTCATTTCTTCACTGATTTGTT							
SpliceSiteFinder-like	[0-100]							
MaxEntScan	[0-16]							
NNSPLICE	[0-1]							
GeneSplicer	[0-15]							
Human Splicing Finder	[0-100]							

REGION 5 – Potential branch-points

	7543-8110	7543-8090	7543-8080	7543-8070	7543-8060	7543-8050	7543-8040
Reference Sequence	CTTCTCATTCTCTAGCAAATCTGCCAAAGTCAGCTGTAGTGCCATACTTGATATGAAGTCTTTGCTGAGA						
Branch Points	[0-100]						

References:

Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 2009;37(9):e67.

Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* 2001;29(5):1185-90.

Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol.* 1997;4(3):311-23.

Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol.* 2004;11(2-3):377-94.