Ejemplo comparación de resultados predictores in sillico

Cambio de estudio BRCA1 c.4484G>T (chr17:41228505 G/T, rs80357389 o NM_007294.3:c.4484G>T)

Exón 13 e intrones adyacentes:

Se ha descrito que este cambio causa la pérdida del exón 14[1].

El cambio se encuentra en la última posición del exón 13 (la **a** en color rojo).

Se va a obtener los resultados que produce analizar esta variable con los diferente predictores y ver cuál de ellos es más preciso.

1. Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C. B., Ficarazzi, F., Barile, M., Varesco, L., Peissel, B., Manoukian, S., & Radice, P. (2013). Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. PloS one, 8(2), e57173. https://doi.org/10.1371/journal.pone.0057173

NetGene2

Donor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' exon intron 3'
328	2	+	0.99	GAGTGGAAAG^GTAAGAAACA H

Donor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' exon intron 3'
139	389	0	-	0.37	GATAATTCAG^GTTAGAATAC

Acceptor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' intron exon 3'
217	0	+	0.07	AACTTCACAG^AAAAGTAGTG
222	2	+	0.07	CACAGAAAAG^TAGTGAATAC
225	2	+	0.17	AGAAAAGTAG^TGAATACCCT
240	2	+	0.18	ACCCTATAAG^CCAGAATCCA
244	0	+	0.18	TATAAGCCAG^AATCCAGAAG
251	1	+	0.18	CAGAATCCAG^AAGGCCTTTC
254	1	+	0.18	AATCCAGAAG^GCCTTTCTGC

Acceptor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' intron exon 3'
294	234	2	-	0.33	TTACTGGTAG^AACTATCTGC

Donor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' exon intron 3'
328	2	+	0.55	GAGTGGAAAT^GTAAGAAACA

Donor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' exon intron 3'
139	389	0	-	0.37	GATAATTCAG^GTTAGAATAC

Acceptor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' intron exon 3'
222	2	+	0.07	CACAGAAAAG^TAGTGAATAC
225	2	+	0.17	AGAAAAGTAG^TGAATACCCT
240	2	+	0.18	ACCCTATAAG^CCAGAATCCA
244	0	+	0.18	TATAAGCCAG^AATCCAGAAG
251	1	+	0.18	CAGAATCCAG^AAGGCCTTTC
254	1	+	0.18	AATCCAGAAG^GCCTTTCTGC

Acceptor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' intron exon 3'
294	234	2	-	0.33	TTACTGGTAG^AACTATCTGC

Desaparece un sitio aceptor en la secuencia mutada (en rojo). Este sitio se encuentra en el interior del exón 13, por lo que, como no se usa en el *splicing* normal (y tiene una *confidence* muy baja) se rechaza que este cambio pueda tener efecto en el *splicing.*

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 89.130.114.18.7237.0 :

Start	End	Score	Exon	Intron
321	335	1.00	tggaaag	**gt**aagaaa

Donor site predictions for 89.130.114.18.7223.0 :

Start	End	Score	Exon	Intron
321	335	0.90	tggaaat	**gt**aagaaa

Acceptor site predictions for 89.130.114.18.7237.0 :

Start	End	Score	Intron	Exon
86	126	0.68	aaaccaatttgtgtatcat	**ag**attgatgcttttgaaaaaaa
400	440	0.87	ttttcaccatactttctcc	**ag**ttttttgcatacaggcattt
414	454	0.79	tctccagttttttgcatac	**ag**gcatttatacacttttattg
440	480	0.82	tatacattttattgctct	**ag**gatacttcttttgtttaatc
468	508	0.97	cttttgtttaatcctatat	**ag**gttttttgaacctataacat

Acceptor site predictions for 89.130.114.18.7223.0 :

Start	End	Score	Intron	Exon
86	126	0.68	aaaccaatttgtgtatcat	**ag**attgatgcttttgaaaaaaa
400	440	0.87	ttttcaccatactttctcc	**ag**ttttttgcatacaggcattt
414	454	0.79	tctccagttttttgcatac	**ag**gcatttatacacttttattg
440	480	0.82	tatacattttattgctct	**ag**gatacttcttttgtttaatc
468	508	0.97	cttttgtttaatcctatat	**ag**gttttttgaacctataacat

No existen diferencias entre ambos resultados, por lo que el cambio no estará afectando al *splicing.*

GENSCAN → no da resultados para este cambio

```
Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------

NO EXONS/GENES PREDICTED IN SEQUENCE
```

```
Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------

NO EXONS/GENES PREDICTED IN SEQUENCE
```

MaxEntScan

```
MAXENT: -16.67  MDD: -18.19     MM: -14.17      WMM: -13.51  MAXENT: -21.86          MM: -22.81              WMM: -21.45
```

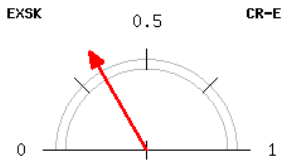
Spliceman

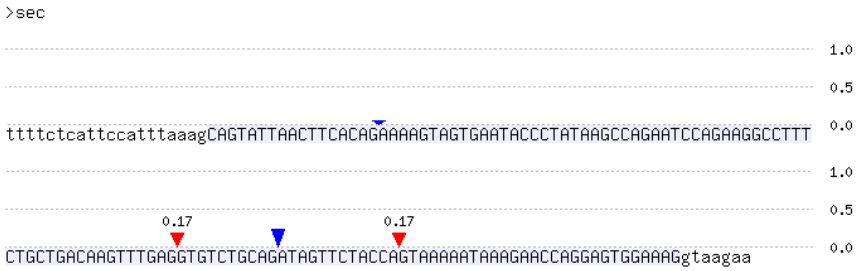
Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ggaaa(g/t)gtaag	aaaggt	aaatgt	30683	78%

En el análisis de la región adyacente al cambio, se obtiene una puntuación muy elevada (78%) para el cambio G>T, por lo que puede estar afectando al *splicing*.

CRYP-SKIP

Exon length (bp)	127
PESS (<=-2.62) density	3.15
NN 5'ss score density	0.05
SF2/ASF score density	8.88
FAS-ESS (hex2) density	3.94
EIE score density	518.26
Probability of cryptic splice site activation (P_{CR-E})	0.33





Parece que hay sitios crípticos de *splicing* dentro del propio exón, pero el cambio de interés (la última **G** en mayúsculas) no lo toma en consideración, por lo que no debe considerar que tenga algún efecto en el *splicing*.

Human Splicing Finder

Type	Interpretation
Alteration of auxiliary sequences	Significant alteration of ESE / ESS motifs ratio (-14)

Algorithm/Matix	position	sequence
ESE_ASF (ESE Site Broken)	chr17:43076551	CTGACAA
ESE_ASFB (ESE Site Broken)	chr17:43076551	CTGACAA
PESE (ESE Site Broken)	chr17:43076551	CTGACAAG
IIE (New ESS Site)	chr17:43076552	GTTGAC
Sironi_motif3 (ESS Site Broken)	chr17:43076552	GCTGACAA
PESE (ESE Site Broken)	chr17:43076552	GCTGACAA
ESE_SRp55 (ESE Site Broken)	chr17:43076553	TGCTGA
IIE (New ESS Site)	chr17:43076553	TGTTGA
Sironi_motif2 (New ESS Site)	chr17:43076553	TGTTGAC
ESE_ASFB (ESE Site Broken)	chr17:43076554	CTGCTGA
PESE (ESE Site Broken)	chr17:43076555	TCTGCTGA
IIE (New ESS Site)	chr17:43076556	TTCTGT
Fas ESS (New ESS Site)	chr17:43076556	TTCTGT
ESE_SC35 (ESE Site Broken)	chr17:43076556	TTCTGCTG
ESE_SRp40 (ESE Site Broken)	chr17:43076557	TTTCTGC
PESS (New ESS Site)	chr17:43076557	TTTCTGTT

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
sec	34	487	taatcattg	-1.686069107	0.526970954357	6	8	17	-0.30962226
sec	34	483	cattgaatt	-0.345039649513	0.52719665272	2	8	17	0.46872033
sec	34	438	tgtttataa	-3.57421817053	0.515011547344	18	9	16	-1.8216801
sec	34	435	ttataaaat	-1.64220173827	0.516279069767	15	9	16	-0.87490004
sec	34	423	gcctgatat	1.33823392681	0.514354066986	3	9	16	1.0510396
sec	34	411	tgtttaaaa	-3.9786620974	0.509852216749	9	7	11	-1.4585996
sec	34	410	gtttaaaaa	-2.02741491785	0.511111111111	8	7	11	-0.6308892
sec	34	392	gtatcatag	-2.5090187538	0.511627906977	9	7	15	-0.84532791
sec	34	384	gattgatgc	1.36336089824	0.514511873351	1	7	15	1.1782097
sec	34	375	ttttgaaaa	-1.87203854993	0.510810810811	25	11	15	-1.6089572
sec	34	366	aaatcagta	-2.95803116112	0.518005540166	16	11	15	-1.462166
sec	34	357	ttctaacct	3.6439487064	0.517045454545	7	11	15	1.692196
sec	34	352	acctgaatt	-0.153490337715	0.515850144092	2	11	15	0.52142466
sec	34	348	gaattatca	-2.09915358828	0.516034985423	30	20	41	-1.7704758
sec	34	345	ttatcacta	-1.81917838995	0.514705882353	27	20	41	-1.4713865
sec	34	339	ctatcagaa	-3.59671435301	0.511976047904	21	20	41	-1.7884676
sec	34	325	cagtaaagt	-0.200262699669	0.525	7	20	41	0.43178655
sec	34	307	ttctcattc	0.636091521069	0.519867549669	1	8	17	0.91381079
sec	34	298	catttaaag	-3.25749960204	0.511945392491	59	10	21	-4.2473121
sec	34	297	atttaaagc	-1.54782725342	0.513698630137	58	10	21	-3.5140299
sec	34	287	gtattaact	-2.47526691872	0.517730496454	48	10	21	-3.242881
sec	34	286	tattaactt	1.20271801391	0.519572953737	47	10	21	-1.7388827
sec	34	281	acttcacag	-1.08027820646	0.514492753623	42	10	21	-2.3179328
sec	34	267	tagtgaata	-1.22495391791	0.530534351145	28	10	21	-1.4832227
sec	34	256	ctataagcc	-0.680878017468	0.529880478088	17	10	21	-0.57412092
sec	34	228	tgctgacaa	2.36883664324	0.533632286996	98	13	22	-4.4966499
sec	34	219	gtttgaggt	-1.89760668064	0.53738317757	89	13	22	-5.5962678
sec	34	192	cagtaaaaa	-0.759149934195	0.545454545455	62	13	22	-3.4388467
sec	34	186	aaataaaga	-1.92512399177	0.558011049724	56	13	22	-3.5115354
sec	34	164	aggtaagaa	-1.949418832	0.610062893082	34	13	22	-2.1116736
sec	34	155	acatcaatg	-3.18887219434	0.626666666667	25	13	22	-2.0219308
sec	34	149	atgtaaaga	-1.64952713808	0.638888888889	19	13	22	-1.0354671
sec	34	131	atctgacat	2.02094704473	0.666666666667	1	13	22	1.5500382
sec	34	123	tctttattt	-4.27070218837	0.661016949153	18	18	31	-1.9074935
sec	34	119	tatttatat	-3.33213260959	0.657894736842	14	18	31	-1.2878141
sec	34	113	tattgaact	0.152040899239	0.666666666667	8	18	31	0.45902748
sec	34	106	ctctgattg	1.70009429717	0.673267326733	1	18	31	1.5103836
sec	34	100	ttgttaatt	-3.45365293947	0.673684210526	2	11	25	-0.62661072
sec	34	99	tgttaattt	-0.0908080195968	0.68085106383	1	11	25	0.75571497
sec	34	90	ttttcacca	0.235616578921	0.658823529412	4	20	35	0.77967487
sec	34	54	catttatac	-2.38440544037	0.65306122449	3	15	26	-0.2685929
sec	34	44	cttttattg	-3.62343691384	0.641025641026	12	14	30	-1.290041
sec	34	18	tgtttaatc	-2.90519198343	0.538461538462	13	0	0	-1.3847129
sec	34	17	gtttaatcc	-0.102649441788	0.583333333333	12	0	0	-0.20959453

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
mut	35	488	taatcattg	-1.686069107	0.530020703934	6	8	17	-0.30863729
mut	35	484	cattgaatt	-0.345039649513	0.530271398747	2	8	17	0.46971339
mut	35	439	tgtttataa	-3.57421817053	0.518433179724	18	9	16	-1.8205751
mut	35	436	ttataaaat	-1.64220173827	0.519721577726	15	9	16	-0.87378821
mut	35	424	gcctgatat	1.33823392681	0.517899761337	3	9	16	1.0521847
mut	35	412	tgtttaaaa	-3.9786620974	0.513513513514	9	7	11	-1.4574171
mut	35	411	gtttaaaaa	-2.02741491785	0.514778325123	8	7	11	-0.62970481
mut	35	393	gtatcatag	-2.5090187538	0.515463917526	9	7	15	-0.84408901
mut	35	385	gattgatgc	1.36336089824	0.518421052632	1	7	15	1.1794723
mut	35	376	ttttgaaaa	-1.87203854993	0.514824797844	25	11	15	-1.6076608
mut	35	367	aaatcagta	-2.95803116112	0.522099447514	16	11	15	-1.4608438
mut	35	358	ttctaacct	3.6439487064	0.521246458924	7	11	15	1.6935528
mut	35	353	acctgaatt	-0.153490337715	0.520114942529	2	11	15	0.52280207
mut	35	349	gaattatca	-2.09915358828	0.520348837209	30	20	41	-1.7690825
mut	35	346	ttatcacta	-1.81917838995	0.519061583578	27	20	41	-1.4699798
mut	35	340	ctatcagaa	-3.59671435301	0.516417910448	21	20	41	-1.787033
mut	35	326	cagtaaagt	-0.200262699669	0.529595015576	7	20	41	0.43327062
mut	35	308	ttctcattc	0.636091521069	0.524752475248	1	8	17	0.91538849
mut	35	299	catttaaag	-3.25749960204	0.517006802721	59	10	21	-4.2456774
mut	35	298	atttaaagc	-1.54782725342	0.518771331058	58	10	21	-3.5123916
mut	35	288	gtattaact	-2.47526691872	0.52296819788	48	10	21	-3.2411894
mut	35	287	tattaactt	1.20271801391	0.524822695035	47	10	21	-1.7371872
mut	35	282	acttcacag	-1.08027820646	0.519855595668	42	10	21	-2.3162007
mut	35	268	tagtgaata	-1.22495391791	0.536121673004	28	10	21	-1.4814181
mut	35	257	ctataagcc	-0.680878017468	0.535714285714	17	10	21	-0.57223677
mut	35	229	tgctgacaa	2.36883664324	0.540178571429	98	13	22	-4.4945357
mut	35	220	gtttgaggt	-1.89760668064	0.544186046512	89	13	22	-5.5940707
mut	35	193	cagtaaaaa	-0.759149934195	0.553191489362	62	13	22	-3.4363479
mut	35	187	aaataaaga	-1.92512399177	0.565934065934	56	13	22	-3.5089765
mut	35	165	atgtaagaa	-2.05259207228	0.6125	34	13	22	-2.1512837
mut	35	156	acatcaatg	-3.18887219434	0.629139072848	25	13	22	-2.0211322
mut	35	150	atgtaaaga	-1.64952713808	0.641379310345	19	13	22	-1.0346628
mut	35	132	atctgacat	2.02094704473	0.669291338583	1	13	22	1.5508859
mut	35	124	tctttattt	-4.27070218837	0.663865546218	18	18	31	-1.9065734
mut	35	120	tatttatat	-3.33213260959	0.660869565217	14	18	31	-1.2868533
mut	35	114	tattgaact	0.152040899239	0.669724770642	8	18	31	0.46001515
mut	35	107	ctctgattg	1.70009429717	0.676470588235	1	18	31	1.5114182
mut	35	101	ttgttaatt	-3.45365293947	0.677083333333	2	11	25	-0.62551292
mut	35	100	tgttaattt	-0.0908080195968	0.684210526316	1	11	25	0.75679998
mut	35	91	ttttcacca	0.235616578921	0.662790697674	4	20	35	0.78095615
mut	35	55	catttatac	-2.38440544037	0.66	3	15	26	-0.26635187
mut	35	45	cttttattg	-3.62343691384	0.65	12	14	30	-1.2871425
mut	35	19	tgtttaatc	-2.90519198343	0.571428571429	14	0	0	-1.4373639
mut	35	18	gtttaatcc	-0.102649441788	0.615384615385	13	0	0	-0.26254121

El único cambio es en el resultado que contiene la mutación. Como está al final del exón y al principio del intrón, puede tener influencia en el *splicing*.

IntSplice

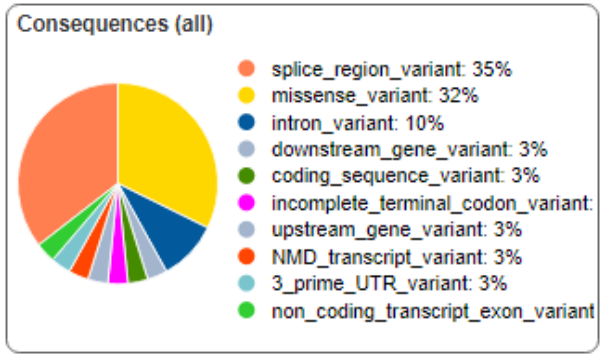
SNV at chr17:41228505 can't be predicted by IntSplice.

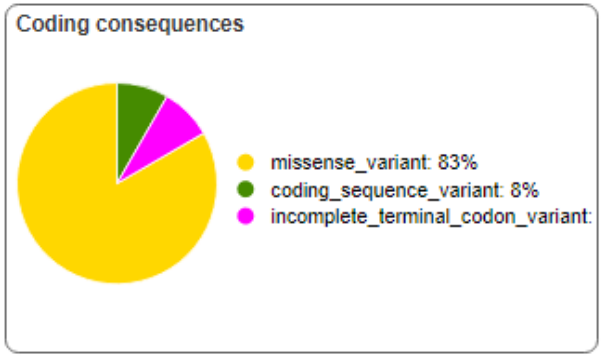
Prediction shows either Abnormal or Normal.

Prediction	Genomic Mutation	Ensembl 64 Transcript ID and Exon No.

Variant Effect Predictor tool

Category	Count
Variants processed	1
Variants filtered out	0
Novel / existing variants	0 (0.0) / 1 (100.0)
Overlapped genes	2
Overlapped transcripts	18
Overlapped regulatory features	0





Se trata de una variante que está afectando al sitio de *splicing*, por lo que va a provocar que se altere el *splicing* normal. Esto se observa en que el 35% de los resultados indican que es una variante en una región de *splcing*, así como el hecho de que es una variante que afecta al 3'UTR y produce NMD, que va a degradar el mRNA mal generado.

Uploaded variant	Location	Allele	Consequence	Symbol	Gene	Feature type	Feature	Biotype
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000352993.7	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000357654.9	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	splice_region_variant, 3_prime_UTR_variant, NMD_transcript_variant	BRCA1	ENSG00000012048	Transcript	ENST00000461221.5	nonsense_mediated_decay
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	downstream_gene_variant	BRCA1	ENSG00000012048	Transcript	ENST00000461574.1	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000468300.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000471181.7	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000478531.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000484087.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	incomplete_terminal_codon_variant, coding_sequence_variant	BRCA1	ENSG00000012048	Transcript	ENST00000487825.5	protein_coding

ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000491747.6	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000493795.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000493919.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	upstream_gene_variant	RPL21P4	ENSG00000240828	Transcript	ENST00000497954.1	processed_pseudogene
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	intron_variant	BRCA1	ENSG00000012048	Transcript	ENST00000586385.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	intron_variant	BRCA1	ENSG00000012048	Transcript	ENST00000591534.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	intron_variant	BRCA1	ENSG00000012048	Transcript	ENST00000591849.5	protein_coding
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	non_coding_transcript_exon_variant	BRCA1	ENSG00000012048	Transcript	ENST00000621897.1	processed_transcript
ENST00000357654.3:c.4484G>T	17:43076488-43076488	A	missense_variant, splice_region_variant	BRCA1	ENSG00000012048	Transcript	ENST00000644379.1	protein_coding

ESEfinder

Cuando se buscan los sitios ESE para la secuencia wt, observamos que el único resultado con la posición de interés que tiene puntuaciones positivas en más de una matriz es GAAAGgt (1.74441, 0.30207, 0.11910, -4.95197). Si buscamos el resultado equivalente para la secuencia mutante, las puntuaciones han pasado a ser negativas en dos de los tres casos positivos (-0.17520,- 1.30179, 0.25754, -2.64314). Por lo tanto, es probable que se esté alterando el ESE interno del exón, produciendo algún tipo de efecto en el *splicing*.

Por otro lado, si buscamos los sitios de *splicing*, solo se obtiene un resultado con puntuación positiva y solo en las matrices 5'SS: ACCAGGAGTGGAAAGgtaagaaacatcaat (11.67400 y 11.30300). Si buscamos el resultado equivalente en la secuencia mutante, las puntuaciones han descendido considerablemente (7.70420 y 7.37980). Por lo tanto, el 5'SS se debilita en la secuencia mutante, lo que puede llevar a cambios en el *splicing.*

EX-SKIP

Seq	PESS (count)	FAS-ESS hex2 (count)	FAS-ESS hex3 (count)	IIE (count)	IIE (sum)	NI-ESS trusted (count)	NI-ESS all (sum)	PESE (count)	RESCUE -ESE (count)	EIE (count)	EIE (sum)	NI-ESE trusted (count)	NI-ESE all (sum)	ESS (total)	ESE (total)	ESS/ESE (ratio)
wt	4	5	2	23	314.3146	17	-24.2516	11	17	45	630.5751	51	63.7701	51	124	0.41
mut	4	5	2	23	314.3146	17	-24.2516	11	17	45	631.5053	51	63.6775	51	124	0.41

Both alleles have a comparable chance of exon skipping.

HOT-SKIP

>wt
tttaatcattgaattccatttggtgctagcatctgtctgttgcattgcttgtgtttataaaattctgcctgatatacttg
tttaaaaaccaatttgtgtatcatagattgatgcttttgaaaaaaatcagtattctaacctgaattatcactatcagaac
aaagcagtaaagtagatttgttttctcattccatttaaagCAGTATTAACTTCACAGAAAAGTAGTGAATACCCTATAAG
CCAGAATCCAGAAGGCCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACCAGGAG
TGGAAAGgtaagaaacatcaatgtaaagatgctgtggtatctgacatctttatttatattgaactctgattgttaatttt
tttcaccatactttctccagttttttgcatacaggcatttatacacttttattgctctaggatacttcttttgtttaatc
ctatataggttttttgaacctataacataagctacaacatgagaaat

Mutation(s) E+76A>G, E+75G>T and E+88A>T have the highest probability of exon skipping.