

Ejemplo comparación de resultados predictores in silico

Cambio de estudio: COL2A1 c.192G>A (chr12: 48393802 G/A, rs121912897 o NM_001844.4:c.192C>T)

Exón 2 e intrones adyacentes

gtaagtcttcccccgcctgctgctgctgcttccatgcgicccctagcatccttc
tcgccggcccgctccagctctggagcccgcggtccgggc.....
.....aatctgtgcaagttgacgcacccctctctcttttctctc
tggtcccccctcttcccttggtctctctgcttctcctccccacccctgggtgcag
AGGAGGCTGGCAGCTGTGTGCAGGATGGGCAGAGGTATAATGATAAGGATGTGTGCAAGC
CGGAGCCCTGCGGATCTGTGTCTGTGACACTGGGACTGTCCTCTGCGACGACATAATCT
GTGAAGACGTGAAAGACTGCCTCAGCCCTGAGATCCCTTCGGAGAGTGTGCCCCATCT
GCCCAACTGACCTCGCCACTGCCAGTG
gttgtaattttattttatttcctgttcaacataaataaaattacttgcaagcactgcaaacac
gctcccatagatgctggtcgtctctgcgaagcagaggggc.....
.....ttttacttgtttgatctaattgagtggttgagtgacaagaa
taggtattttttccatctcaagatctttaaccttcttcttctctatattttttcccttgacag

Se ha descrito que este cambio produce el *skipping* del exón 2 por disrupción del ESE (*exonic splicing enhancer*)¹. Este es un ejemplo de I-NAS. El fenómeno de *nonsense-associated altered splicing*, que se describió como un mecanismo que debería proteger las transcripciones que contienen codones de STOP prematuros (o por sus siglas PTC, *premature termination codon*) del fenómeno de *non-mediated decay*². Durante este proceso, los transcritos que contienen PTC producen un *splicing* alternativo y los fragmentos con codón *nonsense* prematuro se eliminan del RNA maduro.

Non-mediated decay (o por sus siglas en inglés NMD) es un fenómeno post-transcripcional que consiste evitar la producción de proteínas truncadas con efectos deletéreos³.

El cambio de estudio es la segunda **C** que se encuentra marcada en rojo en la segunda fila de nucleótidos del exón. El cambio está descrito como C>T, pero el alelo 'ancestral' es G⁴.

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

1. McAlinden A, Majava M, Bishop PN, Perveen R, Black GC, Pierpont ME, Ala-Kokko L, Männikkö M (2008) Missense and nonsense mutations in the alternatively-spliced exon 2 of COL2A1 cause the ocular variant of Stickler syndrome. Hum Mutat 29(1):83–90

2. Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3(4):285–298

3. Hug, N.; Longman, D.; & Cáceres, J. F. (2015). Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Research, 44(4), 1483–1495.

4. http://grch37.ensembl.org/Homo_sapiens/Variation/Explore?db=core;g=ENSG00000139219;r=12:48366748-48398269;source=dbSNP;t=ENST00000380518;v=rs121912897;vdb=variation;vf=249125965

NetGene2

The sequence: wt has the following composition:

Length: 607 nucleotides.

23.4% A, 25.5% C, 24.1% G, 27.0% T, 0.0% X, 49.6% G+C

Donor splice sites, direct strand

No donor site predictions above threshold.

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
560	48	1	-	0.00	CAACCTGCAA	^	GTGAGAAGTG	
189	419	1	-	0.55	TGCACCAAGG	^	GTGGGGAGGG	

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
70	0	+	0.00	ATATGTCCAG	^	GTGGCCCCAG	
200	1	+	0.33	CTTGGTGCA	^	AGGAGGCTGG	
202	0	+	0.31	TGGTGCA	^	GAGGCTGGCA	
205	0	+	0.20	TGCAGAGGAG	^	GCTGGCAGCT	
213	2	+	0.18	AGGCTGGCAG	^	CTGTGTGCA	
566	0	+	0.00	TCACTTGCA	^	GTTGAAACTG	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
531	77	0	-	0.00	TCTACCCAG	^	CTACCAAGGTC	
491	117	1	-	0.31	TGCTTTGCAG	^	AGACGACCAG	
348	260	2	-	0.07	GGGATCTCAG	^	GGCTGAGGCA	
341	267	0	-	0.17	CAGGGCTGAG	^	GCAGTCTTTC	
337	271	1	-	0.18	GCTGAGGCAG	^	TCTTTCACGT	
319	289	1	-	0.96	GTCTTCACAG	^	ATTATGTCGT	
304	304	1	-	0.20	GTCGTGCA	^	AGGACAGTCC	
302	306	0	-	0.19	CGTCGACAG	^	GACAGTCCCA	
297	311	2	-	0.18	CAGAGGACAG	^	TCCCAGTGTC	
291	317	2	-	0.07	ACAGTCCCAG	^	TGTCACAGAC	

The sequence: mut has the following composition:

Length: 607 nucleotides.

23.4% A, 25.4% C, 24.1% G, 27.2% T, 0.0% X, 49.4% G+C

Donor splice sites, direct strand

No donor site predictions above threshold.

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
560	48	1	-	0.00	CAACCTGCAA	^	GTGAGAAGTG	
189	419	1	-	0.55	TGCACCAAGG	^	GTGGGGAGGG	

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
70	0	+	0.00	ATATGTCCAG	^	GTGGCCCCAG	
200	1	+	0.34	CTTGGTGCA	^	AGGAGGCTGG	
202	0	+	0.31	TGGTGCA	^	GAGGCTGGCA	
205	0	+	0.20	TGCAGAGGAG	^	GCTGGCAGCT	
213	2	+	0.18	AGGCTGGCAG	^	CTGTGTGCA	
566	0	+	0.00	TCACTTGCA	^	GTTGAAACTG	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
531	77	0	-	0.00	TCTACCCAG	^	CTACCAAGGTC	
491	117	1	-	0.31	TGCTTTGCAG	^	AGACGACCAG	
341	267	0	-	0.07	CAGGGCTGAG	^	GCAGTCTTTC	
337	271	1	-	0.17	GCTGAGGCAG	^	TCTTTCACGT	
319	289	1	-	0.95	GTCTTCACAG	^	ATTATGTCGT	
304	304	1	-	0.20	GTCGTGCA	^	AGGACAGTCC	
302	306	0	-	0.19	CGTCACAG	^	GACAGTCCCA	
297	311	2	-	0.19	CAGAGGACAG	^	TCCCAGTGTC	
291	317	2	-	0.18	ACAGTCCCAG	^	TGTCACAGAC	

Al comparar ambas secuencias, comprobamos que la diferencia es que desaparece un sitio *aceptor* en la hebra complementaria. Este sitio se encuentra dentro del exón, por lo que no se utiliza en el *splicing* normal. La mutación, por lo tanto, no tiene ningún efecto en el *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.81.36.659.0 :

Start	End	Score	Exon	Intron
228	242	0.62	ggcagag	gtataatg

Donor site predictions for 85.53.81.36.648.0 :

Start	End	Score	Exon	Intron
228	242	0.62	ggcagag	gtataatg

Acceptor site predictions for 85.53.81.36.659.0 :

Start	End	Score	Intron	Exon
50	90	0.66	ttttaagttctatatgtcc	aggtggccccagcctacattct
546	586	0.94	aaggccacttctcacttgc	aggttgaaactgagtgaacgag

Acceptor site predictions for 85.53.81.36.648.0 :

Start	End	Score	Intron	Exon
50	90	0.66	ttttaagttctatatgtcc	aggtggccccagcctacattct
546	586	0.94	aaggccacttctcacttgc	aggttgaaactgagtgaacgag

Aparecen las mismas predicciones para las secuencias *wild type* (izquierda) y mtada (derecha), por lo que la mutación no tiene ningún efecto en el *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

>sec
TGTCTCTG(C/T)GACGACATAATCTGT MAXENT: -43.57 MDD: -28.15 MM: -28.14 WMM: -24.16

>sec
TGTCTCTG(C/T)GACGACATAATCTGT MAXENT: -18.08 MM: -15.39 WMM: -15.92

Tanto para 5' splice sites (arriba) como 3' splice sites (abajo), los valores son muy negativos, por lo que es poco probable que esté implicado en el *splicing*.

Spliceman

>sec
ACAGATTATGTCGTC(A/G)CAGAGGACA

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
tcgtc(a/g)cagag	cacaga	cgcaga	29741	72%

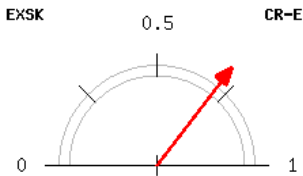
En el análisis de la región adyacente al cambio, se obtiene una puntuación elevada (72%) para el cambio G>A, por lo que puede estar afectando al *splicing*.

CRYP-SKIP

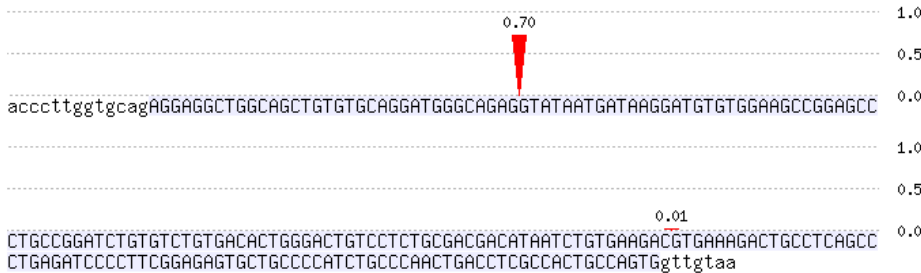
Ni para la secuencia wt (izquierda) ni la mutada (derecha), se indica que hay sitio críptico en la posición de interés.

Results for sequence sec

Exon length (bp)	207
PESS (<=-2.62) density	0.97
NN 5'ss score density	0.30
SF2/ASF score density	21.72
FAS-ESS (hex2) density	3.38
EIE score density	404.81
Probability of cryptic splice site activation (P _{CR-E})	0.71

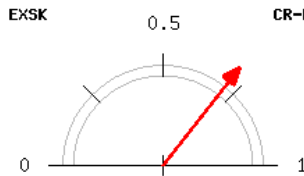


>sec

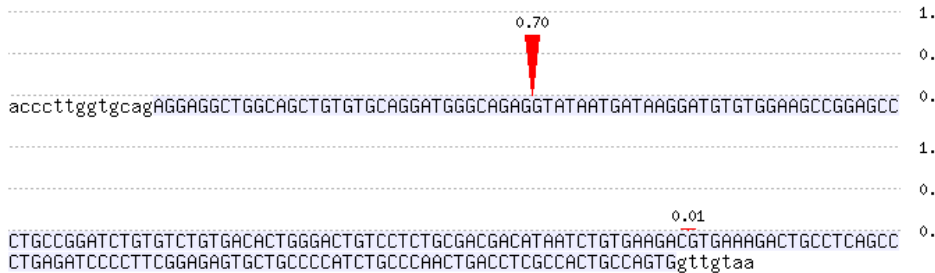


Results for sequence sec

Exon length (bp)	207
PESS (<=-2.62) density	0.97
NN 5'ss score density	0.30
SF2/ASF score density	21.72
FAS-ESS (hex2) density	3.38
EIE score density	404.81
Probability of cryptic splice site activation (P _{CR-E})	0.71



>sec



Human Splicing Finder

Impact Prediction ✕

Format: text Download Copy

Type	Interpretation																																				
Alteration of auxiliary sequences	Significant alteration of ESE / ESS motifs ratio (-7)																																				
<table> <thead> <tr> <th>Algorithm/Matix</th><th>position</th><th>sequence</th></tr> </thead> <tbody> <tr><td>RESCUE ESE (ESE Site Broken)</td><td>chr12:48000019</td><td>CGACGA</td></tr> <tr><td>EIE (ESE Site Broken)</td><td>chr12:48000019</td><td>CGACGA</td></tr> <tr><td>PESE (New ESE Site)</td><td>chr12:48000019</td><td>TGACGACA</td></tr> <tr><td>ESE_SRp55 (ESE Site Broken)</td><td>chr12:48000021</td><td>TGCGAC</td></tr> <tr><td>IIE (New ESS Site)</td><td>chr12:48000021</td><td>TGTGAC</td></tr> <tr><td>Sironi_motif2 (New ESS Site)</td><td>chr12:48000021</td><td>TGTGACG</td></tr> <tr><td>IIE (New ESS Site)</td><td>chr12:48000022</td><td>CTGTGA</td></tr> <tr><td>IIE (New ESS Site)</td><td>chr12:48000023</td><td>TCTGTG</td></tr> <tr><td>Sironi_motif3 (ESS Site Broken)</td><td>chr12:48000023</td><td>TCTGCGAC</td></tr> <tr><td>IIE (New ESS Site)</td><td>chr12:48000024</td><td>CTCTGT</td></tr> <tr><td>ESE_SRp40 (ESE Site Broken)</td><td>chr12:48000025</td><td>CCTCTGC</td></tr> </tbody> </table>	Algorithm/Matix	position	sequence	RESCUE ESE (ESE Site Broken)	chr12:48000019	CGACGA	EIE (ESE Site Broken)	chr12:48000019	CGACGA	PESE (New ESE Site)	chr12:48000019	TGACGACA	ESE_SRp55 (ESE Site Broken)	chr12:48000021	TGCGAC	IIE (New ESS Site)	chr12:48000021	TGTGAC	Sironi_motif2 (New ESS Site)	chr12:48000021	TGTGACG	IIE (New ESS Site)	chr12:48000022	CTGTGA	IIE (New ESS Site)	chr12:48000023	TCTGTG	Sironi_motif3 (ESS Site Broken)	chr12:48000023	TCTGCGAC	IIE (New ESS Site)	chr12:48000024	CTCTGT	ESE_SRp40 (ESE Site Broken)	chr12:48000025	CCTCTGC	
Algorithm/Matix	position	sequence																																			
RESCUE ESE (ESE Site Broken)	chr12:48000019	CGACGA																																			
EIE (ESE Site Broken)	chr12:48000019	CGACGA																																			
PESE (New ESE Site)	chr12:48000019	TGACGACA																																			
ESE_SRp55 (ESE Site Broken)	chr12:48000021	TGCGAC																																			
IIE (New ESS Site)	chr12:48000021	TGTGAC																																			
Sironi_motif2 (New ESS Site)	chr12:48000021	TGTGACG																																			
IIE (New ESS Site)	chr12:48000022	CTGTGA																																			
IIE (New ESS Site)	chr12:48000023	TCTGTG																																			
Sironi_motif3 (ESS Site Broken)	chr12:48000023	TCTGCGAC																																			
IIE (New ESS Site)	chr12:48000024	CTCTGT																																			
ESE_SRp40 (ESE Site Broken)	chr12:48000025	CCTCTGC																																			
New Donor splice site	Activation of a cryptic Donor site. Potential alteration of splicing																																				
<table> <thead> <tr> <th>Algorithm/Matix</th><th>position</th><th>sequences</th><th>variation</th></tr> </thead> <tbody> <tr> <td>HSF Donor site (matrix GT)</td><td>chr12:48000023</td><td>- REF : TCTGCGACG - ALT : TCTGTGACG</td><td>39.22 > 66.36 => 69.2%</td></tr> </tbody> </table>	Algorithm/Matix	position	sequences	variation	HSF Donor site (matrix GT)	chr12:48000023	- REF : TCTGCGACG - ALT : TCTGTGACG	39.22 > 66.36 => 69.2%																													
Algorithm/Matix	position	sequences	variation																																		
HSF Donor site (matrix GT)	chr12:48000023	- REF : TCTGCGACG - ALT : TCTGTGACG	39.22 > 66.36 => 69.2%																																		

La disrupción en un ESE/ESS lleva a alteraciones en el *splicing*, como el *skipping* de un exón completo. En este caso, como el cambio está en el exón 2, podría llevar al *skipping* de dicho exón⁶.

6. Wimmer K, Roca X, Beiglböck H, Callens T, Etzler J, Rao AR, Krainer AR, Fonatsch C, Messiaen L (2007) Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. Hum Mutat 28(6):599–612

SVM-BPfinder → no puede usarse porque es una mutación exónica

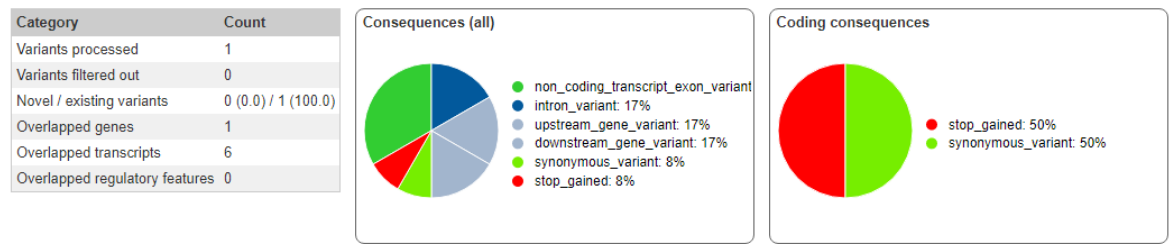
IntSplice

SNV at chr2:48393802 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



Se identifica que tiene efecto en el *splicing*. Si nos fijamos en las consecuencias que produce, la mayoría de los resultados indican que es una *non coding transcript exon variant*, es decir, una variante que se encuentra en un transcrito que no se suele transcribir porque no es el mayoritario. El método de NMD se encarga de degradar estos transcritos⁵, dado que aparece un codón de parada (50% de las consecuencias predichas).

5. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci USA. 2003;100:189–192

Uploaded variant	Location	Allele	Consequence	Symbol	Gene	Feature type	Feature	Biotype	Exon	cDNA position
rs121912897	12:48000019-48000019	A	intron_variant	COL2A1	ENSG00000139219	Transcript	ENST00000337299.7	protein_coding	-	-
rs121912897	12:48000019-48000019	T	intron_variant	COL2A1	ENSG00000139219	Transcript	ENST00000337299.7	protein_coding	-	-
rs121912897	12:48000019-48000019	A	synonymous_variant	COL2A1	ENSG00000139219	Transcript	ENST00000380518.8	protein_coding	2/54	347
rs121912897	12:48000019-48000019	T	stop_gained	COL2A1	ENSG00000139219	Transcript	ENST00000380518.8	protein_coding	2/54	347
rs121912897	12:48000019-48000019	A	downstream_gene_variant	COL2A1	ENSG00000139219	Transcript	ENST00000465743.2	processed_transcript	-	-
rs121912897	12:48000019-48000019	T	downstream_gene_variant	COL2A1	ENSG00000139219	Transcript	ENST00000465743.2	processed_transcript	-	-
rs121912897	12:48000019-48000019	A	upstream_gene_variant	COL2A1	ENSG00000139219	Transcript	ENST00000466884.1	retained_intron	-	-
rs121912897	12:48000019-48000019	T	upstream_gene_variant	COL2A1	ENSG00000139219	Transcript	ENST00000466884.1	retained_intron	-	-
rs121912897	12:48000019-48000019	A	non_coding transcript exon variant	COL2A1	ENSG00000139219	Transcript	ENST00000474996.6	processed_transcript	3/8	430
rs121912897	12:48000019-48000019	T	non_coding transcript exon variant	COL2A1	ENSG00000139219	Transcript	ENST00000474996.6	processed_transcript	3/8	430
rs121912897	12:48000019-48000019	A	non_coding transcript exon variant	COL2A1	ENSG00000139219	Transcript	ENST00000490609.2	retained_intron	2/2	425
rs121912897	12:48000019-48000019	T	non_coding transcript exon variant	COL2A1	ENSG00000139219	Transcript	ENST00000490609.2	retained_intron	2/2	425

ESEfinder

SRSF1 threshold: 1.956	SRSF1 (IgM-BRCA1) threshold: 1.867	SRSF2 threshold: 2.383	SRSF5 threshold: 2.67	SRSF6 threshold: 2.676
Position*/Site/Score	Position*/Site/Score	Position*/Site/Score	Position*/Site/Score	Position*/Site/Score
27 (-71) CTGCCGG 3.50000	27 (-71) CTGCCGG 3.74721	58 (-40) GTCCTCTG 5.04474	47 (-51) ACACTGG 5.02906	37 (-61) TGTGTC 4.58418

* both positions from 5'end (through 1) and 3'end (through -1) are given

Sólo haciendo la búsqueda para el mejor ESE, para ninguna de las matrices corresponde ese fragmento a la región con la mutación.

59 (-39)	TCCTCTG	-3.38453	59 (-39)	TCCTCTG	-2.02341	59 (-39)	TCCTCTGT	-4.17409	59 (-39)	TCCTCTG	-2.21530	59 (-39)	TCCTCT	-3.63835
60 (-38)	CCTCTGT	-1.31690	60 (-38)	CCTCTGT	0.20672	60 (-38)	CCTCTGTG	0.14358	60 (-38)	CCTCTGT	2.29297	60 (-38)	CCTCTG	-2.61782
61 (-37)	CTCTGTG	-0.43523	61 (-37)	CTCTGTG	1.12249	61 (-37)	CTCTGTGA	-1.85197	61 (-37)	CTCTGTG	-3.70164	61 (-37)	CTCTGT	-2.67776
62 (-36)	TCTGTGA	-5.82917	62 (-36)	TCTGTGA	-3.98458	62 (-36)	TCTGTGAC	-5.85897	62 (-36)	TCTGTGA	0.92534	62 (-36)	TCTGTG	0.51690
63 (-35)	CTGTGAC	-2.61903	63 (-35)	CTGTGAC	-0.48415	63 (-35)	CTGTGACG	-1.58186	63 (-35)	CTGTGAC	-3.30280	63 (-35)	CTGTGA	-3.54408
64 (-34)	TGTGACG	-7.85546	64 (-34)	TGTGACG	-5.19152	64 (-34)	TGTGACGA	-2.10737	64 (-34)	TGTGACG	2.42564	64 (-34)	TGTGAC	1.97898
65 (-33)	GTGACGA	3.64303	65 (-33)	GTGACGA	1.69019	65 (-33)	GTGACGAC	-3.40032	65 (-33)	GTGACGA	-4.78852	65 (-33)	GTGACG	-7.01670
66 (-32)	TGACGAC	-5.50068	66 (-32)	TGACGAC	-2.98872	66 (-32)	TGACGACA	0.12252	66 (-32)	TGACGAC	1.53788	66 (-32)	TGACGA	0.51313

Sin embargo, si realizamos la búsqueda para todos los ESE, observamos que el cambio de interés aparece entre los resultados 59 y 66, con valores positivos (en rojo) en algunos de ellos. La mayor puntuación se encuentra en la secuencia GTGACGA con 3.64 de puntuación. Esto quiere decir que es probable que el cambio modifique algún ESE interno, haciendo que se produzcan cambios en el *splicing*.

Si realizamos la búsqueda para los sitios de *splicing*, solo se obtiene un resultado con la posición de interés que tenga puntuación positiva: GCGACGACATAATCTGTGAAGACGTGAAAG, con 0.75290 y 0.57580 para las matrices de 5'SS. Si se busca el fragmento equivalente en la secuencia mutada, se obtienen puntuaciones de 0.12620 y -0.05510. Por lo tanto, está afectando a ese sitio 5'SS se pierde, alterando el *splicing*.

EX-SKIP

EX-SKIP - Results for submitted sequences

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	2	6	4	36	571.7750	12	-17.2533	7	13	72	777.1178	94	123.3321	60	186	0.32
mut	2	6	4	40	633.4044	12	-17.2533	8	12	71	772.4774	93	122.7928	64	184	0.35

Allele mut has a higher chance of exon skipping than allele wt.

Se observa que la mutación produce que exista la posibilidad de *exon skipping* (0.35 frente al 0.32 de la wt).

HOT-SKIP

HOT-SKIP - Results for submitted sequences

```
>mut
aatatttgttgaatacatggaaaaatcaatgatgttttaggaaaaataattttaagttctatatgtccagggtggcccccag
cctacattcttcagcatttgaattctgtcaagttgactgcaacctctctcttttctctctggctccccacccccctctctt
cccttggtctctctgtcttctccctccccaccccttggtgcagAGGAGGCTGGCAGCTGTGTGCAAGATGGGCAAGGTATAA
TGATAAGGATGTGTGGAAGCCGGAGCCCTGCCGGATCTGTCTGTGTGACACTGGGACTGTCTCTGTGACGACATAATCT
GTGAAGACGTGAAAGACTGCCCTCAGCCCTGAGATCCCTTCGGAGAGTGC TGCCCCATCTGCCCAACTGACCTCGCCACT
GCCAGTGgttgtaattttattttctgttcaacataaataaattacttgcaagcactgcaaacacgctcccatagatg
ctggtcgtctctgcaaagcagaggggctagttatccatgggacctggtagctggggtagaaaggaaaggccacttctcac
ttgcagggttgaactgagtgaaacgagcctgagacactagaggggtcc
```

La posición de interés no se encuentra entre las marcadas como probables de que ocurra *splicing*.

Mutation(s) E+81G>T, E+7C>G and E+37A>T have the highest probability of exon skipping.