

## Ejemplo comparación de resultados predictores in silico

Cambio de estudio NPR3 c.892+1G>T (chr5:32724821 G/T, COSV99423083 o NM\_001204375.2: c.892+1G>T)

Exón 2 e intrones adyacentes:

```
ctgccaccttctgattagatgtaggctgdcctatctggacatttttggtcctcaaagtgtt
attccaaggtgtctcatgagctctgtgttcccatgggttggtgatgttgggttccatctcttg
ggtatggaggcgtctgctaggcttcctccatgggtccattttagttaacaagtgattgtga
ggagatacttcaggaccatataagtatcccggttccttgtcagatgtcccttactggtgtc
agcatgctcgaagtgctctgcaaaagggtgctcatgtgtgttgtttgttccctccctag
TGGTGATCATGTGTGCGAGCAGTGACACCATCCGAGCATCATGCTGGTGCGCACAGGC
ATGGCATGACCAAGTGGAGACTACGCTTCTTCAACATGAGCTCTTCAACAGCTCTTCT
ATG
gtaaecttgcttccactttccctccctctgctaggggttccaaagagaggttgtcagatgcc
catgaatggtgggttggaataataatgtggtacataaacaacatggaatactatgcag
ctttataaaaacaatggaatcatgtcttttcagcaacattaatggagctggaggccatga
tcctaaaacaattaatgcaggacagaaaatcaaacactgcagctctcacttacaagtgg
gagctaaacattgagcaaatacgacataaacatgggaacaatagacactggggacttct
```

El cambio se encuentra en la siguiente posición justo después del exón 2 (la g en color granate).

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

## NetGene2

### Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
424		1	+	0.55	TCTTCCTATG	^GTA	ACTCTGC	

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
657	67	0	-	0.00	GCTCCCACTT	^GTA	AGTGAGA	
653	71	0	-	0.00	CCACTTGTA	^GTG	AGAGCAT	
233	491	1	-	0.37	GCTGACACCA	^GTA	AGGGACA	

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
222	0	+		0.30	TCCTTGTCAG	^ATG	TCCCTTA	
300	1	+		0.95	CCTCCCCTAG	^TGG	TGATCAT	
322	2	+		0.34	GTGCGAGCAG	^TGAC	ACCATC	
337	2	+		0.07	CCATCCGGAG	^CAT	CATGCTG	

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
454	270	1	-	0.17	GGAACCCCTAG	^CAG	AGGAGGG	
451	273	1	-	0.17	ACCCTAGCAG	^AGG	AGGGGAA	
449	275	0	-	0.17	CCTAGCAGAG	^GAG	GGGAAAG	
446	278	1	-	0.07	AGCAGAGGAG	^GGG	AAAGTGG	
145	579	0	-	0.20	GACCATGGAG	^GAAC	CTAGCA	
137	587	2	-	0.20	AGGAACCTAG	^CAG	ACGCCTC	
134	590	2	-	0.20	AACCTAGCAG	^ACG	CCTCCAT	
117	607	1	-	0.18	CATAACCAAG	^AGAT	GGAAGC	
115	609	0	-	0.17	TAACCAAGAG	^ATG	GGAAGCCA	

### 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'
657	67	0	-	0.00	GCTCCCACTT	^GTA	AGTGAGA	
653	71	0	-	0.00	CCACTTGTA	^GTG	AGAGCAT	
233	491	1	-	0.37	GCTGACACCA	^GTA	AGGGACA	

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
222	0	+		0.30	TCCTTGTCAG	^ATG	TCCCTTA	
300	1	+		0.95	CCTCCCCTAG	^TGG	TGATCAT	
322	2	+		0.32	GTGCGAGCAG	^TGAC	ACCATC	
337	2	+		0.07	CCATCCGGAG	^CAT	CATGCTG	
457	0	+		0.26	CCTCTGCTAG	^GGT	CCAAGA	

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
454	270	1	-	0.07	GGAACCCCTAG	^CAG	AGGAGGG	
451	273	1	-	0.14	ACCCTAGCAG	^AGG	AGGGGAA	
449	275	0	-	0.07	CCTAGCAGAG	^GAG	GGGAAAG	
446	278	0	-	0.07	AGCAGAGGAG	^GGG	AAAGTGG	
145	579	0	-	0.20	GACCATGGAG	^GAAC	CTAGCA	
137	587	2	-	0.20	AGGAACCTAG	^CAG	ACGCCTC	
134	590	2	-	0.20	AACCTAGCAG	^ACG	CCTCCAT	
117	607	1	-	0.18	CATAACCAAG	^AGAT	GGAAGC	
115	609	0	-	0.17	TAACCAAGAG	^ATG	GGAAGCCA	

Desaparece el sitio *donor* (en azul) de la secuencia WT a la secuencia mutante. Este es el sitio *donor* correspondiente al exón 2 del gen, por lo que su pérdida provocará cambios en el *splicing*, muy probablemente perdiendo el exón completo ya que no se detectan otros sitios *donor*.

Por otro lado, se descubre un sitio *acceptor* (en rojo) críptico en la secuencia mutante que no estaba en la secuencia WT. Este se encuentra después del exón 2, por lo que, si dentro del intrón 2 hubiera un sitio *donor* críptico sin emplear y fueran reconocidos ambos por el *spliceosome*, se podría producir la inclusión de un exón críptico en el mRNA.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
417	431	0.87	tcctatg	gt <sup>a</sup> aactct

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
143	183	0.44	tcctccatgggtccatttgt	ag <sup>a</sup> ttaacaagtgattgtgagga
202	242	0.85	aagtatcccgttccttgtc	ag <sup>a</sup> atgtcccttactgggtgcag
280	320	0.91	tgttgtttgttcctccct	ag <sup>a</sup> tggtgatcatgtgtgcgagc
338	378	0.45	catcatgctgggtggcgac	ag <sup>a</sup> gcatggcatgaccagtggag
437	477	0.97	cactttccctcctctgct	ag <sup>a</sup> ggttccaagagaggttgtca
556	596	0.42	atggaatcatgtcttttgc	ag <sup>a</sup> caacattaatggagctggag

Donor site predictions for mut :

Start	End	Score	Exon	Intron
-------	-----	-------	------	--------

Acceptor site predictions for mut :



Start	End	Score	Intron	Exon
143	183	0.44	tcctccatgggtccatttgt	ag <sup>a</sup> ttaacaagtgattgtgagga
202	242	0.85	aagtatcccgttccttgtc	ag <sup>a</sup> atgtcccttactgggtgcag
280	320	0.91	tgttgtttgttcctccct	ag <sup>a</sup> tggtgatcatgtgtgcgagc
338	378	0.45	catcatgctgggtggcgac	ag <sup>a</sup> gcatggcatgaccagtggag
437	477	0.97	cactttccctcctctgct	ag <sup>a</sup> ggttccaagagaggttgtca
556	596	0.42	atggaatcatgtcttttgc	ag <sup>a</sup> caacattaatggagctggag

Desaparece el sitio *donor* (en azul) de la secuencia WT a la secuencia mutante. Este es el sitio *donor* correspondiente al exón 2 del gen, por lo que su perdida provocará cambios en el *splicing*, muy probablemente perdiendo el exón completo ya que no se detectan otros sitios *donor*.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ctatg(g/t)taact	tatggt	tatggt	29053	68%

## Human Splicing Finder

 Broken WT Donor Site		Alteration of the WT Donor site, most probably affecting splicing	
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr5:32724818	- REF : ATG <b>G</b> TAACT - ALT : ATG <b>T</b> TAACT	84.09 > 56.95 => -32.27%
MaxEnt Donor site	chr5:32724818	- REF : ATG <b>G</b> TAACT - ALT : ATG <b>T</b> TAACT	6.06 > -2.44 => -140.26%
 New Donor splice site		Activation of a cryptic Donor site. Potential alteration of splicing	
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr5:32724817	- REF : TATG <b>G</b> TAAAC - ALT : TATG <b>T</b> TAAAC	42.16 > 69.3 => 64.37%

## SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr			
wt	49	69	tcttcaaca		-1.48624461928	0.703125	20	10	21	-1.0234021		
wt	49	62	cattgagct		0.343769532072	0.736842105263	13	10	21	0.14711284		
wt	49	54	tcttcaaca		-1.48624461928	0.734693877551	5	10	21	-0.063731447		
wt	49	35	<b>tggttaactc</b>		2.47466972148	0.833333333333	1	28	55	<b>2.0889434</b>		
mut	49	69	tcttcaaca		-1.48624461928	0.71875	20	10	21	-1.0183557		
mut	49	62	cattgagct		0.343769532072	0.754385964912	13	10	21	0.15277899		
mut	49	54	tcttcaaca		-1.48624461928	0.755102040816	5	10	21	-0.057140212		
mut	49	36	<b>atgttaact</b>		-3.02643563722	0.806451612903	2	28	55	<b>-0.13697967</b>		
mut	49	35	<b>tggttaactc</b>		2.04706723409	0.833333333333	1	28	55	<b>1.9215168</b>		

Aparece un BP nuevo en la secuencia mutante, pero tiene puntuación positiva, por lo que no se tendrá en cuenta. Además, uno de los sitios con puntuación positiva se ve alterado por la mutación, lo que reduce su *score*, por lo que puede estar debilitándose.

Variant Effect Predictor tool

ENST00000415167.2:c.892+1G>T	<a href="#">5:32724821-32724821</a>	T	splice_donor_variant	NPR3	<a href="#">ENSG00000113389</a>	Transcript	<a href="#">ENST00000265074.13</a>	protein_coding	-	-	-	-	-	-	-	<a href="#">COSV99423083</a>
ENST00000415167.2:c.892+1G>T	<a href="#">5:32724821-32724821</a>	T	splice_donor_variant	NPR3	<a href="#">ENSG00000113389</a>	Transcript	<a href="#">ENST00000326958.5</a>	protein_coding	-	-	-	-	-	-	-	<a href="#">COSV99423083</a>
ENST00000415167.2:c.892+1G>T	<a href="#">5:32724821-32724821</a>	T	splice_donor_variant	NPR3	<a href="#">ENSG00000113389</a>	Transcript	<a href="#">ENST00000415167.2</a>	protein_coding	-	-	-	-	-	-	-	<a href="#">COSV99423083</a>
ENST00000415167.2:c.892+1G>T	<a href="#">5:32724821-32724821</a>	T	splice_donor_variant	NPR3	<a href="#">ENSG00000113389</a>	Transcript	<a href="#">ENST00000434067.6</a>	protein_coding	-	-	-	-	-	-	-	<a href="#">COSV99423083</a>
ENST00000415167.2:c.892+1G>T	<a href="#">5:32724821-32724821</a>	T	splice_donor_variant, non_coding_transcript_variant	NPR3	<a href="#">ENSG00000113389</a>	Transcript	<a href="#">ENST00000506712.1</a>	retained_intron	-	-	-	-	-	-	-	<a href="#">COSV99423083</a>
ENST00000415167.2:c.892+1G>T	<a href="#">5:32724821-32724821</a>	T	splice_donor_variant	NPR3	<a href="#">ENSG00000113389</a>	Transcript	<a href="#">ENST00000509104.5</a>	protein_coding	-	-	-	-	-	-	-	<a href="#">COSV99423083</a>

ESEfinder

En la secuencia mutante aparecen dos predicciones con puntuaciones positivas, 1 para las matrices 5' (409) y otra para las matrices 3' (398):

398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	-10.35130	398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	4.29300	398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	-11.00500	398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	4.25030
409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	6.57500	409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	-7.96370	409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	6.45250	409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	-9.01150

Comparando estas puntuaciones con las de las secuencias mutantes se observan que para 409 la puntuación ha pasado a ser negativa, mientras que para 398 la puntuación se ha reducido ligeramente:

398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	-10.44760	398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	3.58650	398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	-11.22640	398 (-326)	TGAGCTCTTCAACAGCTCTTCCTATGgtaa	3.58710
409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	-10.42950	409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	-10.65630	409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	-10.89530	409 (-315)	ACAGCTCTTCCTATGgtaaactctgcttcca	-11.62270

Por lo tanto se puede concluir que lo más probable es que se esté perdiendo un sitio *donor* en la secuencia mutante, lo que llevará a cambios en el *splicing*.