

## Ejemplo comparación de resultados predictores in silico

Cambio de estudio PAMR1 c.714+5G>C (chr11:35434507 G/C, COSV53509824 o NM\_001001991.3: c.714+5G>C)

Exón 10 e intrones adyacentes:

```
acaggcaagaagaagagaggtctcfaaaagattagacaacttgcccaaggttctattgatag
taagtgcctgggatcagaactcaggtttcatttgactgcaaactcctagactttatagcca
gccatgtcagggtgggttgattaatatatttaaattgcagaagaaaactgggttacttatctt
agtggtcacatgcccatacagttcaccttttcagatttggctctctgggttaagtgcaca
catctccctttggatggcaaaagtctgagtttatcttactaagctgggtactttctccatag
TCTGCGGAAAAATTGAGAATCACTGCTCCAAAGACCCAAGGTTGCGCTGGCCGTGGC
AGGCAGCCATCTACAGGAGGACAGCGGGGTGCATGACGGCAGCCTACACAAGGGAGCGT
GGTTCCTAGTCTGCAGCGGTGCCCTGGTGAATGAGCGCACTGTGGTGGTGGCTGCCCACT
GTGTTACTGACCTGGGGAAGGTCAACATGATCAAGACAGCAGACCTGAAAGTTGTTTTGG
GGAAATCTACCGGGATGATGACCGCCATGAGAAGACCATCCAGAGCCTACAG
gtagagaggggcttgcacttggaagctggggctcttggctcaaaaaaacccaagcagggg
agagcagtgattataccaagctctgtgcaccttagcatgtcagcagccagcccagagcctc
gcataatagcagcgttcattatttcataattgagatactagatgggggtacctagaaaggct
gtggagtcctttttccaccttgattattttttaataagatagaggcaaaggcatctatc
actgtggtcatcaaaagttaatacacctaagagtaaaatgtactctctcctagtagcacatg
```

El cambio se encuentra en la quinta posición justo después del exón 10 (la g en color naranja).

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'
60	0	+	0.00	TCTATTGATA	^GTAAGTGCCT		
594	0	+	1.00	GAGCCTACAG	^GTAAGAGAGG	H	
604	1	+	0.00	GTAAGAGAGG	^GCTTGCACTT		

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
491	403	0	-	0.82	CCTTCCCAG	^GTCAGTAACA		

### Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
215	1	+	0.16	ACCTTTCCAG	^ATTTGTGCTC		
362	0	+	0.23	GCCGTGGCAG	^GCAGCCATCT		
366	1	+	0.07	TGGCAGGCAG	^CCATCTACAG		
376	2	+	0.17	CCATCTACAG	^GAGGACCGAG		
379	2	+	0.17	TCTACAGGAG	^GACCGAGCGG		
385	2	+	0.17	GGAGGACCG	^CGGGGTGCAT		
403	2	+	0.17	ATGACGGCAG	^CCTACACAAG		
413	0	+	0.17	CCTACACAAG	^GGAGCGTGGT		
429	1	+	0.07	TGGTTCCTAG	^TCTGCAGCGG		
436	2	+	0.14	TAGTCTGCAG	^CGGTGCCCTG		
687	2	+	0.33	GTCCCCTTAG	^CATGTCAGCA		

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
619	275	2	-	0.07	AGAGCCCCAG	^CTTCCAAGTG		
611	283	1	-	0.17	AGCTTCCAAG	^TGCAAGCCCT		
605	289	1	-	0.17	CAAGTGCAAG	^CCCTCTCTTA		
588	306	0	-	0.97	TTACCTGTAG	^GCTCTGGATG		
371	523	0	-	0.30	CCTCCTGTAG	^ATGGCTGCCT		
302	592	2	-	0.16	TTTCCCGCAG	^ACTATGGAGA		

### Donor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
60	0	+	0.00	TCTATTGATA	^GTAAGTGCCT		
594	0	+	0.94	GAGCCTACAG	^GTAACAGAGG	H	

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
491	403	0	-	0.81	CCTTCCCAG	^GTCAGTAACA		

### Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
215	1	+	0.16	ACCTTTCCAG	^ATTTGTGCTC		
362	0	+	0.23	GCCGTGGCAG	^GCAGCCATCT		
366	1	+	0.07	TGGCAGGCAG	^CCATCTACAG		
376	2	+	0.17	CCATCTACAG	^GAGGACCGAG		
379	2	+	0.17	TCTACAGGAG	^GACCGAGCGG		
385	2	+	0.17	GGAGGACCG	^CGGGGTGCAT		
403	2	+	0.17	ATGACGGCAG	^CCTACACAAG		
413	0	+	0.17	CCTACACAAG	^GGAGCGTGGT		
429	1	+	0.07	TGGTTCCTAG	^TCTGCAGCGG		
436	2	+	0.14	TAGTCTGCAG	^CGGTGCCCTG		
687	2	+	0.33	GTCCCCTTAG	^CATGTCAGCA		

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
627	267	0	-	0.14	TTTGAGCCAG	^AGCCCCAGCT		
625	269	2	-	0.07	TGAGCCAGAG	^CCCCAGCTTC		
619	275	2	-	0.17	AGAGCCCCAG	^CTTCCAAGTG		
611	283	1	-	0.17	AGCTTCCAAG	^TGCAAGCCCT		
605	289	1	-	0.17	CAAGTGCAAG	^CCCTCTGTTA		
588	306	0	-	0.82	TTACCTGTAG	^GCTCTGGATG		
371	523	0	-	0.30	CCTCCTGTAG	^ATGGCTGCCT		
302	592	2	-	0.16	TTTCCCGCAG	^ACTATGGAGA		

Pérdida de sitio *donor* (en rojo) en la secuencia mutante. Este sitio se encuentra al inicio del intrón 10, pero no es el sitio *donor* del exón, por lo tanto es probable que no tenga efecto en el *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.2.148.387935.0 :

Start	End	Score	Exon	Intron
53	67	0.98	attgata	gt aagtgc
279	293	0.52	taagctg	gt actttc
587	601	1.00	cctacag	gt aagaga

Donor site predictions for 10.42.0.139.388361.0 :

Start	End	Score	Exon	Intron
53	67	0.98	attgata	gt aagtgc
279	293	0.52	taagctg	gt actttc
587	601	0.95	cctacag	gt aacaga

Acceptor site predictions for 10.42.2.148.387935.0 :

Start	End	Score	Intron	Exon
195	235	0.66	ccatacagttcacctttcc	ag atttgtgctctctggttaag
280	320	0.96	aagctgggtactttctccat	ag tctgcgggaaaattgagaac
416	456	0.43	agcgtgggttcctagtctgc	ag cgggtgccctggatgagc
667	707	0.78	taccaagctctgtcccctt	ag catgtcagcagccagcccag

Acceptor site predictions for 10.42.0.139.388361.0 :


Start	End	Score	Intron	Exon
195	235	0.66	ccatacagttcacctttcc	ag atttgtgctctctggttaag
280	320	0.96	aagctgggtactttctccat	ag tctgcgggaaaattgagaac
416	456	0.43	agcgtgggttcctagtctgc	ag cgggtgccctggatgagc
667	707	0.78	taccaagctctgtcccctt	ag catgtcagcagccagcccag

La única diferencia que hay entre las dos secuencias es que el último sitio *donor* de los detectados tiene un cambio a consecuencia de la presencia de la mutación (la quinta posición en el intrón en vez de ser una G es una C en la secuencia mutante). Sin embargo el sitio *donor* sigue siendo predicho, por lo tanto es muy probable que no esté afectado el *splicing*.


Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ggtaa(g/c)agagg	ggtaag	ggtaac	26333	51%

## 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)	chr11:35434514	- REF : CAGGTAA <b>G</b> A - ALT : CAGGTAA <b>C</b> A	97.33 > 87.31 => -10.29%

 New Acceptor splice site		Activation of a cryptic Acceptor site. Potential alteration of splicing	
Algorithm/Matix	position	sequences	variation
HSF Acceptor site (matrix AG)	chr11:35434516	- REF : TACAGGTAA <b>G</b> AGAG - ALT : TACAGGTAA <b>C</b> AGAG	67.57 > 75.21 => 11.31%

## SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	24	68	ggatgatga		-0.708808573926	0.365079365079	46	9	14	-2.5391452	
wt	24	65	tgatgaccg		2.0115288642	0.366666666667	43	9	14	-1.2835969	
wt	24	56	ggatgagaa		-2.79573592634	0.372549019608	34	9	14	-2.5942835	
wt	24	30	aggtaagag		-1.36718598693	0.36	8	9	14	-0.39323533	
mut	24	68	ggatgatga		-0.708808573926	0.380952380952	46	9	14	-2.5340187	
mut	24	65	tgatgaccg		2.0115288642	0.383333333333	43	9	14	-1.278214	
mut	24	56	ggatgagaa		-2.79573592634	0.392156862745	34	9	14	-2.5879508	
mut	24	30	aggtaacag		0.999371603929	0.4	8	9	14	0.54630263	

La diferencia entre ambas es el cambio en la región que incluye la posición de interés donde el *score* pasa de ser negativo a positivo, por lo que puede estar activando un BP en la secuencia mutante y estar afectando el *splicing*.

## Variant Effect Predictor tool

ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	splice_region_variant, intron_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000527605.5</a>	protein_coding	-	-	<a href="#">COSV53509824</a>
ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	downstream_gene_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000531219.1</a>	retained_intron	-	-	<a href="#">COSV53509824</a>
ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	splice_region_variant, intron_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000611014.4</a>	protein_coding	-	-	<a href="#">COSV53509824</a>
ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	splice_region_variant, intron_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000615849.4</a>	protein_coding	-	-	<a href="#">COSV53509824</a>
ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	splice_region_variant, intron_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000619888.5</a>	protein_coding	-	-	<a href="#">COSV53509824</a>
ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	splice_region_variant, intron_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000621476.4</a>	protein_coding	-	-	<a href="#">COSV53509824</a>
ENST00000611014.4:c.714+5G>C	<a href="#">11:35434507-35434507</a>	G	splice_region_variant, intron_variant	PAMR1	<a href="#">ENSG00000149090</a>	Transcript	<a href="#">ENST00000622144.4</a>	protein_coding	-	-	<a href="#">COSV53509824</a>

## ESEfinder

Solo se observa una predicción con puntuaciones positivas y se dan en los dos tipos de matrices:

579 (-315)	ATCCAGAGCCTACAGgtaagagagggcttg	13.46420	579 (-315)	ATCCAGAGCCTACAGgtaagagagggcttg	4.98990	579 (-315)	ATCCAGAGCCTACAGgtaagagagggcttg	13.52280	579 (-315)	ATCCAGAGCCTACAGgtaagagagggcttg	4.93130
---------------	--------------------------------	----------	---------------	--------------------------------	---------	---------------	--------------------------------	----------	---------------	--------------------------------	---------

Cuando se compara con las puntuaciones de la predicción equivalente en la secuencia mutante, se observa que las puntuaciones para las matrices 5' han descendido, mientras que para 3' han aumentado ligeramente. Por lo tanto, lo más probable es que se está debilitando un sitio *donor* que podría estar afectando al *splicing*, aunque tampoco se puede descartar la activación de un *acceptor*.

579 (-315)	ATCCAGAGCCTACAGgtaacagagggcttg	9.66130	579 (-315)	ATCCAGAGCCTACAGgtaacagagggcttg	5.35660	579 (-315)	ATCCAGAGCCTACAGgtaacagagggcttg	9.74300	579 (-315)	ATCCAGAGCCTACAGgtaacagagggcttg	5.24720
---------------	--------------------------------	---------	---------------	--------------------------------	---------	---------------	--------------------------------	---------	---------------	--------------------------------	---------