

Ejemplo comparación de resultados predictores in silico

Cambio de estudio TSPAN17 c.747+1G>T (chr 5:176656817 G/T, COSV53781727 o NM_130465.5: c.747+1G>T)

Exón 7 e intrones adyacentes:

```
ccagagatttgggtgctcatgagacagtccggctctggaggagtggagggagaaaagacgggta
agcccttagagaaaagcaatatgaacagccatgtgacagctctcagcctgagagccttgaaga
aaataaggaggtgataggagagaagctgggcaggcaggagggggcgtgcattctgtgctg
agaccatgagaaggagccggccttctggagggaactgtggggagcaggtactgcagggtt
tagaccctgggggtgggcgtgagccttggcaggcctgggagggtgggagccggcctgaaggggc
aggtaggctgagcctgggtgcctgcgtgtccccag
GAGCTGAGCAGCAGGGCTTCATCCACACCAAAGGCTGCGTGGCCAGTTTGAGAAGTGG
CTCAGGACAACTGATTGTGTTGGCGGAGTCTTCATGGGCATCGCCCTCTCCAG
gtacccttgtggccccaagtgccctccctcttgcgggggcgcccctcctctccctcac
ctgtcctctgtcttacag
```

El cambio se encuentra en siguiente posición justo después del exón 7 (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'
58		0	+	0.00	GAGAAAGACG	^	GTAAGCCCTT	
229		1	+	0.39	TGGGGAGCAG	^	GTAAGCCCTT	
452		0	+	0.00	CCTCCTCCAG	^	GTAAGCCCTT	

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
511		19		-	-	0.00	CAGAGGACAG	^	GTGAGGGGAG	
455		75		1	-	0.00	GGCCACAAGG	^	GTAAGCCCTT	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
306		2	+	0.17	GGGCAGGTAG	^	GCTGAGCCTG	
334		0	+	0.83	GTGTCCCCAG	^	GAGCTGGAGC	
337		0	+	0.76	TCCCCAGGAG	^	CTGGAGCAGC	
343		0	+	0.55	GGAGCTGGAG	^	CAGCAGGGCT	
346		0	+	0.44	GCTGGAGCAG	^	CAGGGCTTCA	
349		0	+	0.42	GGAGCAGCAG	^	GGCTTCATCC	
368		1	+	0.17	CACACCAAG	^	GCTGCGTGGG	
451		1	+	0.00	CCTCCTCCAG	^	GTAAGCCCTT	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
427		103		1	-	0.19	GCCCATGAAG	^	ACTCCCGCCA	
407		123		0	-	0.33	CCACAATCAG	^	GTTGTCTCTG	
395		135		0	-	0.19	TGTCCTGCAG	^	CCACTTCTCA	
292		238		1	-	0.27	TGCCCTTCAG	^	GCCGGCTCCC	
216		314		0	-	0.26	CTCCCCACAG	^	TCCCTCCACA	
112		418		0	-	0.26	TTCTTCCAAG	^	CTCTCAGCTG	
65		465		0	-	0.00	TTTCTCTAAG	^	GGCTTACCGT	
33		497		0	-	0.00	ACTCCTCCAG	^	AGCCGACTGT	

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
58		0	+	0.00	GAGAAAGACG	^	GTAAGCCCTT	
229		1	+	0.39	TGGGGAGCAG	^	GTAAGCCCTT	

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
511		19		-	-	0.00	CAGAGGACAG	^	GTGAGGGGAG	
455		75		1	-	0.00	GGCCACAAGG	^	GTAAGCCCTT	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
306		2	+	0.17	GGGCAGGTAG	^	GCTGAGCCTG	
334		0	+	0.83	GTGTCCCCAG	^	GAGCTGGAGC	
337		0	+	0.76	TCCCCAGGAG	^	CTGGAGCAGC	
343		0	+	0.55	GGAGCTGGAG	^	CAGCAGGGCT	
346		0	+	0.44	GCTGGAGCAG	^	CAGGGCTTCA	
349		0	+	0.34	GGAGCAGCAG	^	GGCTTCATCC	
368		1	+	0.18	CACACCAAG	^	GCTGCGTGGG	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
427		103		1	-	0.18	GCCCATGAAG	^	ACTCCCGCCA	
407		123		0	-	0.44	CCACAATCAG	^	GTTGTCTCTG	
395		135		0	-	0.34	TGTCCTGCAG	^	CCACTTCTCA	
292		238		1	-	0.27	TGCCCTTCAG	^	GCCGGCTCCC	
216		314		0	-	0.26	CTCCCCACAG	^	TCCCTCCACA	
112		418		0	-	0.26	TTCTTCCAAG	^	CTCTCAGCTG	
65		465		0	-	0.00	TTTCTCTAAG	^	GGCTTACCGT	
33		497		0	-	0.00	ACTCCTCCAG	^	AGCCGACTGT	

Pérdida sitio *donor* (en rojo) de la secuencia WT a la mutante. A pesar de su nivel de confianza, este sitio coincide con el sitio *donor* del exón, por lo tanto se va a producir un cambio el *splicing*, lo más probable la inclusión de un intrón porque los otros *donor* detectados están en el intrón antes del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.1.119.358175.0 :

Start	End	Score	Exon	Intron
51	65	1.00	aaagacg	gt aagccc
222	236	0.44	ggagcag	gt actgca
296	310	0.90	agggcag	gt aggctg
445	459	0.92	cctccag	gt accctt

Donor site predictions for 10.42.2.148.358997.0 :

Start	End	Score	Exon	Intron
51	65	1.00	aaagacg	gt aagccc
222	236	0.44	ggagcag	gt actgca
296	310	0.90	agggcag	gt aggctg

Acceptor site predictions for 10.42.1.119.358175.0 :

Start	End	Score	Intron	Exon
314	354	0.73	ctggcgctgcgtgtcccc	aggagctggagcagcagggtt
431	471	0.81	atgggcatcgccctcctcc	aggtacccttgtggccccacgt

Acceptor site predictions for 10.42.2.148.358997.0 :


Start	End	Score	Intron	Exon
314	354	0.73	ctggcgctgcgtgtcccc	aggagctggagcagcagggtt

Se pierde un sitio *donor* (en rojo) y un sitio *acceptor* (en verde) en la secuencia mutante. El sitio *donor* coincide con el sitio *donor* del exón, mientras que el sitio *acceptor* que está prediciendo coincide con el final del exón (realmente detecta un sitio *donor* pero lo indica como *acceptor*), por lo que no se tendrá en cuenta. Por lo tanto, la pérdida del sitio *donor* conducirá muy probablemente a la pérdida del exón completo, ya que los sitios *donor* anteriores están por delante del exón.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
tccag(g/t)tacc	tccagg	tccagt	28496	64%

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)	chr5:176656814	- REF : CAGGTACCC - ALT : CAGTTACCC	79.33 > 52.19 => -34.21%
MaxEnt Donor site	chr5:176656814	- REF : CAGGTACCC - ALT : CAGTTACCC	8.63 > 0.13 => -98.49%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr	
wt	78	99	tcttcatgg		-0.525384456446	0.734042553191	7	12	23	0.20431541
wt	78	31	ccctcactc		2.51211956019	0.884615384615	1	26	50	2.0735901
wt	78	20	ccctcacct		2.96283869771	0.866666666667	1	15	29	2.0486383
mut	78	99	tcttcatgg		-0.525384456446	0.744680851064	7	24	36	0.32885725
mut	78	76	cagttaccc		0.0320011763507	0.774647887324	17	16	28	-0.15073141
mut	78	31	ccctcactc		2.51211956019	0.884615384615	1	26	50	2.0735901
mut	78	20	ccctcacct		2.96283869771	0.866666666667	1	15	29	2.0486383

Variant Effect Predictor tool

ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	missense_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000298564.14	protein_coding	4/4	573	424	142	V/L	GTA/TTA	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000310032.12	protein_coding	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant, NMD_transcript_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000503030.1	nonsense_mediated_decay	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000503045.5	protein_coding	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000504168.5	protein_coding	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000507471.1	protein_coding	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000508164.6	protein_coding	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant, non_coding_transcript_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000514705.5	retained_intron	-	-	-	-	-	-	COSV53781727
ENST00000508164.5:c.747+1G>T	5:176656817-176656817	T	splice_donor_variant	TSPAN17	ENSG00000048140	Transcript	ENST00000515708.5	protein_coding	-	-	-	-	-	-	COSV53781727

ESEfinder

Se obtienen 2 predicciones con puntuaciones positivas para las matrices 5' (445 y 437) y una predicción con puntuación positiva para las matrices 3' (437):

437 (-93)	ATCGCCCTCCTCCAGgtacaccttggtggccc	6.86590	437 (-93)	ATCGCCCTCCTCCAGgtacaccttggtggccc	11.59600	437 (-93)	ATCGCCCTCCTCCAGgtacaccttggtggccc	7.04710	437 (-93)	ATCGCCCTCCTCCAGgtacaccttggtggccc	11.58810
445 (-85)	CCTCCAGgtacaccttggtggcccccacgtgcc	0.69810	445 (-85)	CCTCCAGgtacaccttggtggcccccacgtgcc	-16.72410	445 (-85)	CCTCCAGgtacaccttggtggcccccacgtgcc	0.69420	445 (-85)	CCTCCAGgtacaccttggtggcccccacgtgcc	-18.64220

Comparando las puntuaciones para las predicciones equivalentes en la secuencia mutante, se observa que la puntuación para la predicción 445 se mantiene parecida. Para la predicción 437 las puntuaciones de las matrices 5' descienden drásticamente hasta una puntuación negativa, mientras que las puntuaciones para la matriz 3' aumentan.

437 (-93)	ATCGCCCTCCTCCAGttacaccttggtggccc	-10.13860	437 (-93)	ATCGCCCTCCTCCAGttacaccttggtggccc	8.90340	437 (-93)	ATCGCCCTCCTCCAGttacaccttggtggccc	-10.30070	437 (-93)	ATCGCCCTCCTCCAGttacaccttggtggccc	8.97690
445 (-85)	CCTCCAGttacaccttggtggcccccacgtgcc	0.47870	445 (-85)	CCTCCAGttacaccttggtggcccccacgtgcc	-15.06090	445 (-85)	CCTCCAGttacaccttggtggcccccacgtgcc	0.48270	445 (-85)	CCTCCAGttacaccttggtggcccccacgtgcc	-17.01920

Por lo tanto lo más probable es que se esté perdiendo un sitio *donor*, lo cual afectara al *splicing*.