

Ejemplo comparación de resultados predictores in silico

Cambio de estudio NXF2 c.233+2T>C (chr23:102317154 T/C, COSV100364093 o NM_022053.4: c.233+2T>C)

Exón 4 e intrones adyacentes:

```
ccccctccaggtgtgtggctgggtgtctagtctaaccatggacacgcgcacctctgttct
ttataattcctgctccatttctcaggaggttcaacacagttgcttcctctcccaagattc
ctaccctgattgtgcccactgtgagataaggctcttgtatctggacaggcagaggagttt
gttttctggggccttatgtggcgggttggcactttgctagccctaccatggagttagctca
cgacgtgggtacgtggtaagtggaccattgatgtgtctcaaacttgctctttttgttctcag
TTGCAGATTGCCATGACCATGGTAGCACTTTTCAAGGAAGAAAGAAAGGTGGGAGTTCTT
TCCGGGATAATTGTGACAAGAGGAGCTGTCATTATGAACATGGTGGTATGAGCGCCCGC
CTTCACACTGCCAGGAGAAATGATGGAAGCGTGGAGATGAGGATGTCCACAAGGACCAAC
AACTAAGACA
gttaagtgaccaggcagcctggtttgcacgtagcagcccccgggactgtgcaaccctttca
ttctctgtggctcttctcttttctctctcattagagaactgaagaaatgctggaagtggaat
agtggctgagcagtcctaattgtagccctggcgtcagtgagtgagcatgtataggagac
ttctcttagatttaatggataccgccttctctctcttccccacag
```

El cambio se encuentra en la segunda posición justo después del exón 4 (la **t** 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'
11	-	+		0.00	CCCC	TTCCAG	^GTGTGTGGCT	
248	0	+		0.76	TCAC	GACGTG	^GTACGTGGTA	
255	1	+		0.83	GTGG	TACGTG	^GTAAGTGGAC	
349	0	+		0.31	AGAA	GAAAG	^GTGGGAGTTC	
403	1	+		0.51	TATG	AACATG	^GTGGGTATGA	
491	2	+		0.82	AACT	AAGACA	^GTAAGTGACC	
647	1	+		0.00	CCTG	GCGTCA	^GTGAGTGGAG	

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
225		492		1	-	0.45	TAACT	CCATG	^GTAGGGCTAG	
124		593		0	-	0.47	CACA	ATCAGG	^GTAGGAATCT	
51		666		1	-	0.00	AAGA	ACAGAG	^GTGCGCGTGT	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
300	0	+		0.43	TTGT	TCTCAG	^TTGCAGAATG	
306	0	+		0.77	TCAG	TTCAG	^AATGCCATGA	
583	2	+		0.38	CTCT	CATTAG	^AGAACTGACG	
585	1	+		0.16	CTCA	TAGAG	^AACTGACGAA	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
428		289		1	-	0.33	CTCCT	GGCAG	^TGTGAAGGCG	
161		556		0	-	0.25	GCCT	GTCCAG	^ATACAAGAGC	
139		578		1	-	0.43	TATCT	CACAG	^GTGGCACAAT	

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
11	-	+		0.00	CCCC	TTCCAG	^GTGTGTGGCT	
248	0	+		0.76	TCAC	GACGTG	^GTACGTGGTA	
255	1	+		0.83	GTGG	TACGTG	^GTAAGTGGAC	
403	1	+		0.47	TATG	AACATG	^GTGGGTATGA	
647	1	+		0.00	CCTG	GCGTCA	^GTGAGTGGAG	

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
225		492		1	-	0.45	TAACT	CCATG	^GTAGGGCTAG	
124		593		0	-	0.47	CACA	ATCAGG	^GTAGGAATCT	
51		666		1	-	0.00	AAGA	ACAGAG	^GTGCGCGTGT	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
300	0	+		0.43	TTGT	TCTCAG	^TTGCAGAATG	
306	0	+		0.77	TCAG	TTCAG	^AATGCCATGA	
583	2	+		0.38	CTCT	CATTAG	^AGAACTGACG	
585	1	+		0.16	CTCA	TAGAG	^AACTGACGAA	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
428		289		1	-	0.30	CTCCT	GGCAG	^TGTGAAGGCG	
161		556		0	-	0.25	GCCT	GTCCAG	^ATACAAGAGC	
139		578		1	-	0.43	TATCT	CACAG	^GTGGCACAAT	

Desaparecen dos sitios *donor* que se predicen para la secuencia WT. El primero (en azul) corresponde a una región que se encuentra dentro del exón, por lo que no se tendrá en cuenta. Sin embargo el otro sitio (en rojo) corresponde con el sitio *donor* de exón, por lo que se producirá un cambio en el *splicing*. Hay dos posibilidades: o que se produzca el *skipping* del exón completo o que se emplee como sitio *donor* alguna de las otras predicciones. Si se tiene en cuenta el sitio *acceptor* que se emplea para este exón, solo podrían utilizarse los dos últimos sitios predichos, provocando la pérdida de los últimos 88 nucleótidos del exón o bien la inclusión de 156 nucleótidos, respectivamente.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
4	18	0.87	cttcag	gtgtgtgg
248	262	0.99	gtacgtg	gt aagtgg
342	356	0.76	aagaaag	gtgggagt
396	410	0.78	gaacatg	gtgggtat
484	498	0.97	taagaca	gt aagtga

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
65	105	0.78	aattcctgctccatttctc	aggaggttcaacacagttgctt
96	136	0.82	acagttgcttctctccca	agattcctaccctgattgtgcc
280	320	0.97	aacttgctttttgttctc	agttgcagaatgccatgaccat
563	603	0.96	ttccttttcttctctcatt	agagaactgacgaatgctggaa

Donor site predictions for mut :

Start	End	Score	Exon	Intron
4	18	0.87	cttcag	gtgtgtgg
248	262	0.99	gtacgtg	gt aagtgg
342	356	0.76	aagaaag	gtgggagt
396	410	0.78	gaacatg	gtgggtat

Acceptor site predictions for mut :


Start	End	Score	Intron	Exon
65	105	0.78	aattcctgctccatttctc	aggaggttcaacacagttgctt
96	136	0.82	acagttgcttctctccca	agattcctaccctgattgtgcc
280	320	0.97	aacttgctttttgttctc	agttgcagaatgccatgaccat
563	603	0.96	ttccttttcttctctcatt	agagaactgacgaatgctggaa

Se pierde uno de los sitios *donor* que se detectan para la secuencia WT este corresponde con el sitio *donor* del exón por lo que se va a producir un camión el *splicing*. Se produciría la pérdida del exón completo o, si se emple el sitio *acceptor* del exón, solo se podrían utilizar los dos últimos *donor* predichos, produciendo la escisión de 146 y 88 nucleótidos, respectivamente.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gacag(t/c)aagtg	acagta	acagca	25742	47%

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)	chrX:102317150	- REF : ACAGTAAAGT - ALT : ACAGCAAGT	83.61 > 56.46 => -32.47%
MaxEnt Donor site	chrX:102317150	- REF : ACAGTAAAGT - ALT : ACAGCAAGT	9.49 > 1.74 => -81.66%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	18	39	aactaagac		0.254338351937		0.441176470588	34	0	0	-1.5082925
wt	18	31	cagtaagtg		0.199767267621		0.5 26	0	0	-1.0042748	
wt	18	27	aagtgacca		1.35364911614		0.545454545455	22	0	0	-0.28460174
mut	18	39	aactaagac		0.254338351937		0.441176470588	34	0	0	-1.5082925
mut	18	27	aagtgacca		1.35364911614		0.545454545455	22	0	0	-0.28460174

Desaparece uno de los sitios BP por la presencia de la mutación. Como este tenía puntuación negativa, no se va a tener en cuenta.

Variant Effect Predictor tool

ENST00000625106.3:c.233+2T>C	X:102317154-102317154	C	 splice_donor_variant	NXF2	ENSG00000269405	Transcript	ENST00000604790.2	protein_coding	-	-	-	-	-	-	COSV100364093
ENST00000625106.3:c.233+2T>C	X:102317154-102317154	C	 splice_donor_variant	NXF2	ENSG00000269405	Transcript	ENST00000625106.4	protein_coding	-	-	-	-	-	-	COSV100364093

ESEfinder

Se obtienen 2 predicciones con puntuaciones positivas para las matrices 5'SS (476 y 480).

476 (-241)	CCAACAACCTAAGACAgtaagtaccaggca	8.48420	476 (-241)	CCAACAACCTAAGACAgtaagtaccaggca	-33.34060	476 (-241)	CCAACAACCTAAGACAgtaagtaccaggca	8.58850	476 (-241)	CCAACAACCTAAGACAgtaagtaccaggca	-35.24500
477 (-240)	CAACAACCTAAGACAgtaagtaccaggcag	-25.82390	477 (-240)	CAACAACCTAAGACAgtaagtaccaggcag	-3.39940	477 (-240)	CAACAACCTAAGACAgtaagtaccaggcag	-22.45990	477 (-240)	CAACAACCTAAGACAgtaagtaccaggcag	-4.03250
478 (-239)	AACAACCTAAGACAgtaagtaccaggcagc	-24.70460	478 (-239)	AACAACCTAAGACAgtaagtaccaggcagc	-35.62590	478 (-239)	AACAACCTAAGACAgtaagtaccaggcagc	-20.85500	478 (-239)	AACAACCTAAGACAgtaagtaccaggcagc	-35.40000
479 (-238)	ACAACCTAAGACAgtaagtaccaggcagcc	-31.77580	479 (-238)	ACAACCTAAGACAgtaagtaccaggcagcc	-34.51870	479 (-238)	ACAACCTAAGACAgtaagtaccaggcagcc	-29.86380	479 (-238)	ACAACCTAAGACAgtaagtaccaggcagcc	-39.76400
480 (-237)	CAACTAAGACAgtaagtaccaggcagcct	3.06180	480 (-237)	CAACTAAGACAgtaagtaccaggcagcct	-14.29960	480 (-237)	CAACTAAGACAgtaagtaccaggcagcct	3.50450	480 (-237)	CAACTAAGACAgtaagtaccaggcagcct	-18.59110

Si se comparan estas puntuaciones con las de la secuencia mutante se comprueba que para 476 el descenso en las puntuaciones es muy brusco, mientras que para 480 aumenta en 2 puntos.

476 (-241)	CCAACAACCTAAGACAgcaagtaccaggca	1.54810	476 (-241)	CCAACAACCTAAGACAgcaagtaccaggca	-33.80400	476 (-241)	CCAACAACCTAAGACAgcaagtaccaggca	1.47840	476 (-241)	CCAACAACCTAAGACAgcaagtaccaggca	-35.71050
477 (-240)	CAACAACCTAAGACAgcaagtaccaggcag	-19.58480	477 (-240)	CAACAACCTAAGACAgcaagtaccaggcag	-2.48720	477 (-240)	CAACAACCTAAGACAgcaagtaccaggcag	-15.88020	477 (-240)	CAACAACCTAAGACAgcaagtaccaggcag	-3.22780
478 (-239)	AACAACCTAAGACAgcaagtaccaggcagc	-25.60380	478 (-239)	AACAACCTAAGACAgcaagtaccaggcagc	-33.82370	478 (-239)	AACAACCTAAGACAgcaagtaccaggcagc	-21.75390	478 (-239)	AACAACCTAAGACAgcaagtaccaggcagc	-32.63500
479 (-238)	ACAACCTAAGACAgcaagtaccaggcagcc	-31.63870	479 (-238)	ACAACCTAAGACAgcaagtaccaggcagcc	-39.08660	479 (-238)	ACAACCTAAGACAgcaagtaccaggcagcc	-29.76890	479 (-238)	ACAACCTAAGACAgcaagtaccaggcagcc	-41.39010
480 (-237)	CAACTAAGACAgcaagtaccaggcagcct	5.21260	480 (-237)	CAACTAAGACAgcaagtaccaggcagcct	-12.71130	480 (-237)	CAACTAAGACAgcaagtaccaggcagcct	5.60640	480 (-237)	CAACTAAGACAgcaagtaccaggcagcct	-16.97740

Por lo tanto dónde mayor cambio a es en 476, lo que probablemente esté implicando un sitio *donor* más débil.