

## Ejemplo comparación de resultados predictores in silico

Cambio de estudio NIN c.1546-2A>G (chr 14:50766398 A/G, COSV55382420 o NM\_016350.5: c.1546-2A>G)

Exón 14 e intrones adyacentes:

```
ccagtttggtgttgttcgtgcttacatgtcacatttcctatcactcaagagactgggtga  
ataccagaaactgggcctttgctgtgttgggtcaaagaagggaactcgggaaaaatgaca  
gctttgccccttttcccttctag  
TTTGGTGACCTCGATCCTAGCAGTGCTGAGTTCTTCCTGCAAGAAGAGAGACTGACACAG  
ATGAGAAATGAATATGAGCGGCAGTGCAGG  
gtaggtagaacaagagaactgagtaagtgcaggagtgcctgggttcagaccagcttcc  
ccgttcctcctagctctgtgaccttaagcaagtcacttcactgtgggtadtccatttcctc  
atgcataaaatggtaaaaataaaagtagccaccacacaaaaatgttgttagcaaaaattccatc
```

El cambio se encuentra en la segunda posición justo antes del exón 14 (la **a** 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'
117		1	+	0.34	AAGAGACTGG	^	GTGAATACCA	
294		0	+	0.87	GCAGTGCAGG	^	GTAAGGTAGAC	
316		1	+	0.83	AGAGAACTGA	^	GTAAGTGCAG	

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
84	390	1	-	0.00	AATGTGACAT	^	GTAAGCACGA	

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
203		0	+	0.97	CCCCTTCTAG	^	TTTGGTGACC	
223		2	+	0.18	TCGATCCTAG	^	CAGTGCTGAG	
226		2	+	0.17	ATCCTAGCAG	^	TGCTGAGTTC	

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
255	219	0	-	0.19	TCTGTGTCAG	^	TCTCTCTTCT	
240	234	0	-	0.97	CTTCTTGACAG	^	GAAGAACTCA	
236	238	1	-	0.14	TTGCAGGAAG	^	AACTCAGCAC	
131	343	0	-	0.07	AAAGGCCAG	^	TTTCTGGTAT	
113	361	0	-	0.17	ATTCACCCAG	^	TCTCTTGAGT	

### Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
117		1	+	0.32	AAGAGACTGG	^	GTGAATACCA	
294		0	+	0.89	GCAGTGCAGG	^	GTAAGGTAGAC	
316		1	+	0.83	AGAGAACTGA	^	GTAAGTGCAG	

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
84	390	1	-	0.00	AATGTGACAT	^	GTAAGCACGA	

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
223		2	+	0.17	TCGATCCTAG	^	CAGTGCTGAG	
226		2	+	0.17	ATCCTAGCAG	^	TGCTGAGTTC	

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
255	219	0	-	0.20	TCTGTGTCAG	^	TCTCTCTTCT	
240	234	0	-	0.97	CTTCTTGACAG	^	GAAGAACTCA	
236	238	1	-	0.18	TTGCAGGAAG	^	AACTCAGCAC	
113	361	0	-	0.14	ATTCACCCAG	^	TCTCTTGAGT	

Desaparece uno de los sitios *acceptores* de *splicing* (en rojo). Este coincide con el sitio *acceptor* del exón por lo que el *splicing* va a ser alterado. Existen dos posibilidades: bien que se produzca la pérdida del exón completo o que se utilice uno de los otros dos aceptores predichos (siempre que se siga empleando el sitio *donor* natural del exón), lo que conduciría a la pérdida de 20 o 23 nt del exón, respectivamente.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.2.148.425733.0 :

Start	End	Score	Exon	Intron
110	124	0.59	agactgg	gtgaatac
287	301	0.95	gtgcagg	gtaggtag
309	323	0.99	gaactga	gt aagtgc
419	433	0.87	taaaatg	gtaaaaat

Donor site predictions for 10.42.3.123.425745.0 :

Start	End	Score	Exon	Intron
110	124	0.59	agactgg	gtgaatac
287	301	0.95	gtgcagg	gtaggtag
309	323	0.99	gaactga	gt aagtgc
419	433	0.87	taaaatg	gtaaaaat

Acceptor site predictions for 10.42.2.148.425733.0 :

Start	End	Score	Intron	Exon
183	223	0.98	tttgcccttttcccccttct	agtttggtgacctcgatcctag
344	384	0.91	cccagcttccccgttcct	agctctgtgaccctaagcaagt

Acceptor site predictions for 10.42.3.123.425745.0 :


Start	End	Score	Intron	Exon
344	384	0.91	cccagcttccccgttcct	agctctgtgaccctaagcaagt

Se produce la pérdida de un sitio *acceptor* (en rojo) en la secuencia mutante. Este coincide con el sitio *acceptor* del exón, por lo que se va a ver alterado el *splicing*, probablemente con la pérdida de exón completo ya que el otro sitio *acceptor* que se predice se encuentra una vez pasada la región del exón.


Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
cttct(a/g)gtttg	cttcta	cttctg	29279	69%

## Human Splicing Finder

 Broken WT Acceptor Site	Alteration of the WT Acceptor site, most probably affecting splicing		
Algorithm/Matix	position	sequences	variation
HSF Acceptor site (matrix AG)	chr14:50766408	- REF : TTCCCCTTCTAGTT - ALT : TTCCCCTTCTGGTT	85.92 > 58.05 => -32.44%
MaxEnt Acceptor site	chr14:50766416	- REF : TTGCCCTTTTCCCCTTCTAGTTT - ALT : TTGCCCTTTTCCCCTTCTGGTTT	10.84 > 2.88 => -73.43%

 New Donor splice site	Activation of a cryptic Donor site. Potential alteration of splicing		
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr14:50766400	- REF : CTAGTTTGG - ALT : CTGGTTTGG	68.18 > 78.26 => 14.78%

## SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	23	46	tggtgacct		2.63890969484	0.512195121951	1	10	15	1.6768991	
wt	23	25	tgctgagtt		0.891296046665	0.45	2	10	21	0.9651353	
mut	23	46	tggtgacct		2.63890969484	0.512195121951	1	10	15	1.6768991	
mut	23	25	tgctgagtt		0.891296046665	0.45	2	10	21	0.9651353	

## Variant Effect Predictor tool

ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000245441.9</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000324330.13</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000382041.7</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000382043.8</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000389869.7</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000453196.5</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	downstream_gene_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000453401.6</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000476352.5</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000530853.5</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000530997.7</a>	protein_coding	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	intron_variant, NMD_transcript_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000673657.1</a>	nonsense_mediated_decay	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	upstream_gene_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000673819.1</a>	processed_transcript	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	splice_acceptor_variant, NMD_transcript_variant	NIN	<a href="#">ENSG00000100503</a>	Transcript	<a href="#">ENST00000674030.1</a>	nonsense_mediated_decay	-	-	<a href="#">COSV55382420</a>
ENST00000453196.5:c.1546-2A>G	<a href="#">14:50766398-50766398</a>	C	regulatory_region_variant	-	-	RegulatoryFeature	<a href="#">ENSR00000496841</a>	promoter_flanking_region	-	-	<a href="#">COSV55382420</a>

## ESEfinder

Se obtienen dos resultados con puntuaciones positivas uno para las matrices 3' (189) y otro para las matrices 5' (193):

189 (-285)	cttttcccccttctagTTTGGTGACCTCGAT	-13.08620	189 (-285)	cttttcccccttctagTTTGGTGACCTCGAT	10.86160	189 (-285)	cttttcccccttctagTTTGGTGACCTCGAT	-13.47780	189 (-285)	cttttcccccttctagTTTGGTGACCTCGAT	10.87220
193 (-281)	tcccccttctagTTTGGTGACCTCGATCCTA	1.98210	193 (-281)	tcccccttctagTTTGGTGACCTCGATCCTA	-4.81360	193 (-281)	tcccccttctagTTTGGTGACCTCGATCCTA	1.68380	193 (-281)	tcccccttctagTTTGGTGACCTCGATCCTA	-5.73230

Sin embargo, cuando comparamos las puntuaciones con las que se obtienen la secuencia mutante, se observa que para 193 aumentan ligeramente en 5', mientras que para 189 pasan a ser puntuaciones negativas:

189 (-285)	cttttcccccttctggTTTGGTGACCTCGAT	-15.39170	189 (-285)	cttttcccccttctggTTTGGTGACCTCGAT	-5.71770	189 (-285)	cttttcccccttctggTTTGGTGACCTCGAT	-15.84040	189 (-285)	cttttcccccttctggTTTGGTGACCTCGAT	-3.66510
193 (-281)	tcccccttctggTTTGGTGACCTCGATCCTA	2.07540	193 (-281)	tcccccttctggTTTGGTGACCTCGATCCTA	-4.90080	193 (-281)	tcccccttctggTTTGGTGACCTCGATCCTA	1.75510	193 (-281)	tcccccttctggTTTGGTGACCTCGATCCTA	-5.88860

Además se observa que el resultado 188 pasa de tener puntuaciones negativas en la secuencia *wild type* a puntuaciones positivas para todas las matrices en la secuencia mutante:

188 (-286)	ccttttcccccttctagTTTGGTGACCTCGA	-2.07270	188 (-286)	ccttttcccccttctagTTTGGTGACCTCGA	-11.82260	188 (-286)	ccttttcccccttctagTTTGGTGACCTCGA	-2.25750	188 (-286)	ccttttcccccttctagTTTGGTGACCTCGA	-16.14630
188 (-286)	ccttttcccccttctggTTTGGTGACCTCGA	1.31470	188 (-286)	ccttttcccccttctggTTTGGTGACCTCGA	2.16230	188 (-286)	ccttttcccccttctggTTTGGTGACCTCGA	1.03140	188 (-286)	ccttttcccccttctggTTTGGTGACCTCGA	1.20170

Por lo tanto se está activando un sitio *donor* y se está perdiendo o se altere un sitio *acceptor* en la secuencia mutante.