

## Ejemplo comparación de resultados predictores in silico

Cambio de estudio SEMA4A c.685+1G>A (chr1:156160560 G/A, COSV100740529 o NM\_022367.4: c.685+1G>A)

Exón 7 e intrones adyacentes:

```
aagtgaattctacttggcaagtggttcaagctagggtatgtgacgaagtagccacagaaaggc
tggaaaagtgccccaaacacggagggccacacaggatggccatttcctgtggagagaggaa
gctaaagaggggtgttttagtgaggagggagctggcctttgaggtccaggaaggtgc
agaactgatgcagggcccccgagactgatattggatgccagccagtcagaggcaggtct
gtggagggcagaggaaccctccacccatcagttctctcttctcctctcctccaccctag
ATGGGATGCTCTATTCTGGTACTATGAACAACCTCCTGGCAGTGAGCCCATCCTGATGC
GCACACTGGGATCCAGCCTGTCCTCAAGACCGACAACCTCCTCGCTGGCTGCATC
gtaaggacctgaccctcgctggcctcctttcctgagtcctggtgaccttgctaattcact
caactttcatatttgcggaaggtacaaatgtgtccattactgttaggcgcagggggtatatgg
cagggaagaaggcaaagctccggtctctcttgaagctttcgtcttatgtaaacaattaaa
gaagtaccagtcattccccacccacatcgctaccctcgacaagctgtgggacccgagga
```

El cambio se encuentra en la primera 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'
33		0	+	0.00		GTT	CAGCTAG^GTATGTGACG	
418		1	+	0.82		TGGCTGCATC^GTAAGGACCT		

### Donor splice sites, complement strand

No donor site predictions above threshold.

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
272		0	+	0.18		ACCCCATCAG^TTCTCTCTTC		
300		1	+	0.85		TCCACCCTAG^ATGGGATGCT		

### Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
239		419		1	-	0.71		CCCTCCACAG^ACCTGCCTCT		
106		552		0	-	0.71		CTCTCCACAG^GAAATGGCCA		

### Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
33		0	+	0.00		GTT	CAGCTAG^GTATGTGACG	

### Donor splice sites, complement strand

No donor site predictions above threshold.

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
272		0	+	0.18		ACCCCATCAG^TTCTCTCTTC		
300		1	+	0.85		TCCACCCTAG^ATGGGATGCT		

### Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
239		419		1	-	0.71		CCCTCCACAG^ACCTGCCTCT		
106		552		0	-	0.71		CTCTCCACAG^GAAATGGCCA		

Se pierde uno de los sitios *donor* (en azul) en la secuencia mutante. Este coincide con el sitio *donor* del exón, por lo que va a tener efecto en el *splicing*, lo más probable que perdiéndose el exón completo por la poca confianza del otro sitio y porque este se encuentra por delante del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
26	40	0.99	cagctag	gtatgtga
411	425	0.92	ctgcatc	gt aaggac
490	504	0.63	gcggaag	gtacaatg

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
280	320	0.99	ttctcctctcctccaccct	agatgggatgctctattctggt

Donor site predictions for mut :

Start	End	Score	Exon	Intron
26	40	0.99	cagctag	gtatgtga
490	504	0.63	gcggaag	gtacaatg

Acceptor site predictions for mut :


Start	End	Score	Intron	Exon
280	320	0.99	ttctcctctcctccaccct	agatgggatgctctattctggt

Se pierde uno de los sitios *donor* en la secuencia mutante. Este coincide con el sitio *donor* del exón, por lo tanto se van a producir cambios en el *splicing*, bien perdiendo el exón completo o bien produciéndose la inclusión de 71 pb (si se usa el siguiente *donor* y asumiendo que se emplea el mismo sitio *acceptor*).

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gcatc(g/a)taagg	tcgtaa	tcataa	29076	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)	chr1:156160557	- REF : ATC <b>G</b> TAAGG - ALT : ATC <b>A</b> TAAGG	80.35 > 53.21 => -33.78%
MaxEnt Donor site	chr1:156160557	- REF : ATC <b>G</b> TAAGG - ALT : ATC <b>A</b> TAAGG	5.84 > -2.34 => -140.07%

## SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	17	50	tcgtaagga		-1.81720786416	0.666666666667	4	26	39	0.015693581	
wt	17	42	acctgaccc		3.00926732845	0.72972972973	1	21	36	2.0878017	
wt	17	19	tcctgagtc		1.03287939407	0.642857142857	14	0	0	0.12764647	
mut	17	53	gcatacataa		-2.88769463265	0.645833333333	7	26	39	-0.60007612	
mut	17	50	tcataagga		-2.61788702629	0.666666666667	4	26	39	-0.29781017	
mut	17	42	acctgaccc		3.00926732845	0.72972972973	1	21	36	2.0878017	
mut	17	19	tcctgagtc		1.03287939407	0.642857142857	14	0	0	0.12764647	

La mutación produce que la puntuación del BP pase de positiva a negativa, por lo que puede que se esté perdiendo un BP.

## Variant Effect Predictor tool

ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000355014.6</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000368282.1</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000368284.5</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000368285.8</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000368286.6</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000414683.5</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000435124.5</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000438830.5</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	upstream_gene_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000462892.1</a>	retained_intron	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant, non_coding_transcript_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000466698.1</a>	retained_intron	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	upstream_gene_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000469065.1</a>	retained_intron	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	downstream_gene_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000470306.5</a>	retained_intron	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	downstream_gene_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000485575.1</a>	processed_transcript	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	splice_donor_variant, non_coding_transcript_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000487358.5</a>	processed_transcript	-	-	-	-	-	-	<a href="#">COSV100740529</a>
ENST00000368285.7:c.685+1G>A	<a href="#">1:156160560-156160560</a>	A	downstream_gene_variant	SEMA4A	<a href="#">ENSG00000196189</a>	Transcript	<a href="#">ENST00000633494.1</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV100740529</a>

## ESEfinder

Se encuentran dos predicciones con puntuación positiva para la secuencia WT: una para las matrices 5'SS (403) y otra para las matrices 3' SS (408):

403 (-255)	TCCGCTGGCTGCATCgtaaggacctgaccc	5.94710	403 (-255)	TCCGCTGGCTGCATCgtaaggacctgaccc	-18.82590	403 (-255)	TCCGCTGGCTGCATCgtaaggacctgaccc	6.08560	403 (-255)	TCCGCTGGCTGCATCgtaaggacctgaccc	-19.58700
408 (-250)	TGGCTGCATCgtaaggacctgacccccgct	-14.74160	408 (-250)	TGGCTGCATCgtaaggacctgacccccgct	2.20310	408 (-250)	TGGCTGCATCgtaaggacctgacccccgct	-10.92180	408 (-250)	TGGCTGCATCgtaaggacctgacccccgct	2.02860

Si se comparan estas puntuaciones con las equivalentes en la secuencia mutante si observa que para 403 las puntuaciones han pasado negativas, mientras que para 408 han aumentado muy ligeramente:

403 (-255)	TCCGCTGGCTGCATCataaggacctgacct	-5.03540	403 (-255)	TCCGCTGGCTGCATCataaggacctgacct	-20.15590	403 (-255)	TCCGCTGGCTGCATCataaggacctgacct	-5.04270	403 (-255)	TCCGCTGGCTGCATCataaggacctgacct	-20.83740
408 (-250)	TGGCTGCATCataaggacctgacccccgct	-14.51350	408 (-250)	TGGCTGCATCataaggacctgacccccgct	2.34820	408 (-250)	TGGCTGCATCataaggacctgacccccgct	-10.65540	408 (-250)	TGGCTGCATCataaggacctgacccccgct	2.12070

Por lo tanto, lo más probable es que se esté perdiendo un sitio de *donor* en la secuencia mutante.