

## Ejemplo comparación de resultados predictores in silico

Cambio de estudio KLHL2 c.1480+2T>C (chr 4:165313368 T/C, COSV56979927 o NM\_007246.4: c.1480+2T>C)

Exón 12 e intrones adyacentes:

```
ccttttttcgattttcctaagataacttttggtattttaattctgtcttaaggatggtgtt
cacttatcaaatagtgttaacctaaagatataaatgagtggtgctattttatcagattattga
tgaaagtataaaattaacatcatcagctataacctgcagatcttcataacatgatttgat
taccocatctgtcaccatttaggcaagaacctaatatatttcataaaaaatcagcagcaatt
taagggggaaactctgctgccatgaaggaaaatatattaataattttctggcttgaaaaatt
agtgttttttggttggttggttttttaataaattttggctttctatgtgatttttatgttag
GTTTGCTCTATGCTGTAGGAGGTATGATGGAGCATCACGTCAAGTGTCTTAGCACAGTAG
AATGCTATTAATGCTACAACAAATGAGTGGACCTATATAGCAGAAATGAGCACCAGGCCGA
GTGGAGCAG
gtacatgtgaactctgtttttagcaactgaagcacaataatgatggaaaataggttaaagtaa
aatgatctcagtggcatcactttacatttatctgaaaatatatgatatgctgattttagtgg
cactgctaaatgtcttttgaaattgtgcaggtaaatgggagtaattacaagtaattttact
ttaaattttaatgtctaagtgccatagacatttctgtttcatcttatatttcaagtgtt
agcacctccttttagttttttgttttaggtgcttattggggaaacttgcttaataatcaagt
cttttcatgtatatcttcttaaaatttctgtgctttaatatgctgttaattttactatggata
```

El cambio se encuentra en la segunda posición justo después del exón 12 (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'
490		1	+	0.66	AGTGGAGCAG	^	GTACATGTGA	
639		0	+	0.47	AATTGTGCAG	^	GTAATGGGAG	

### 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'
51		2	+	0.00	TGCTTACAG	^	GATGTTGTTC	
360		1	+	0.49	TTATGTGTAG	^	GTTTGCTCTA	
378		1	+	0.29	TATGCTGTAG	^	GAGGTTATGA	
757		2	+	0.00	TTGTTTTTAG	^	TGCTTATTGG	

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
579	271	0	-	0.25	ATATTTTCAG	^	ATAATGTAAA	
452	398	1	-	0.27	TGCTATATAG	^	GTCCACTCAT	

### Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
639		0	+	0.47	AATTGTGCAG	^	GTAATGGGAG	

### 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'
51		2	+	0.00	TGCTTACAG	^	GATGTTGTTC	
360		1	+	0.43	TTATGTGTAG	^	GTTTGCTCTA	
378		1	+	0.27	TATGCTGTAG	^	GAGGTTATGA	
757		2	+	0.00	TTGTTTTTAG	^	TGCTTATTGG	

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
579	271	0	-	0.25	ATATTTTCAG	^	ATAATGTAAA	
452	398	1	-	0.28	TGCTATATAG	^	GTCCACTCAT	

Pérdida sitio *donor* (en rojo) de la secuencia WT a la mutante. Este sitio coincide con el sitio *donor* del exón, por lo que, si no se emplea el otro sitio *donor* predicho, se estará perdiendo el exón completo. En caso de que se usara el otro sitio, el exón pasaría de 129 a 278 bp.

### Splice Site Prediction by Neural Network (NNSplice)

#### Donor site predictions for 10.42.1.119.338819.0 :

Start	End	Score	Exon	Intron
393	407	0.48	gcatcac	gtcagtgt
483	497	0.93	ggagcag	gtacatgt
632	646	0.84	tgtgcag	gtaatggg

#### Donor site predictions for 10.42.0.139.339335.0 :

Start	End	Score	Exon	Intron
393	407	0.48	gcatcac	gtcagtgt
632	646	0.84	tgtgcag	gtaatggg

#### Acceptor site predictions for 10.42.1.119.338819.0 :

Start	End	Score	Intron	Exon
31	71	0.91	gtattttaattctgtctttac	aggatgttgttcacttatcaaa
340	380	0.88	tctatgtgattttatgtgt	aggtttgctctatgctgtagga
358	398	0.77	taggtttgctctatgctgt	aggaggttatgatggagcatca
618	658	0.69	aatgtcttttgaattgtgc	aggtaatgggagtaattacaaa
723	763	0.58	agttgttagcactcctttt	agtttttgttttagtgctta
737	777	0.98	cttttagtttttgttttt	agtgtttattggggaacttgct

#### Acceptor site predictions for 10.42.0.139.339335.0 :


Start	End	Score	Intron	Exon
31	71	0.91	gtattttaattctgtctttac	aggatgttgttcacttatcaaa
340	380	0.88	tctatgtgattttatgtgt	aggtttgctctatgctgtagga
358	398	0.77	taggtttgctctatgctgt	aggaggttatgatggagcatca
618	658	0.69	aatgtcttttgaattgtgc	aggtaatgggagtaattacaaa
723	763	0.58	agttgttagcactcctttt	agtttttgttttagtgctta
737	777	0.98	cttttagtttttgttttt	agtgtttattggggaacttgct

Pérdida sitio *donor* (en rojo) de la secuencia WT a la mutante. Este sitio coincide con el sitio *donor* del exón, por lo que o se está perdiendo el exón completo o, si se empleara el siguiente sitio predicho, el exón pasaría de 129 a 278 bp. Si se empleara el *donor* anterior, el exón pasaría de 129 a 39 bp.

### Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gcagg(t/c)acatg	caggta	caggca	33553	60%

## 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)	chr4:165313364	- REF : CAGGTACAT - ALT : CAGGCACAT	82.71 > 55.57 => -32.81%
MaxEnt Donor site	chr4:165313364	- REF : CAGGTACAT - ALT : CAGGCACAT	7.79 > 0.04 => -99.49%

## SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr			
wt	22	67	aaatgagtg		-0.985617479759	0.387096774194	62	0	0	-3.7836129		
wt	22	44	aaatgagca		-1.12366269252	0.410256410256	39	0	0	-2.3743229		
mut	22	67	aaatgagtg		-0.985617479759	0.387096774194	62	0	0	-3.7836129		
mut	22	44	aaatgagca		-1.12366269252	0.410256410256	39	0	0	-2.3743229		

## Variant Effect Predictor tool

ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000226725.11</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000421009.6</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	downstream_gene_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000506541.5</a>	processed_transcript	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000506761.1</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant, NMD_transcript_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000506824.5</a>	nonsense_mediated_decay	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant, non_coding_transcript_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000509028.1</a>	processed_transcript	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000514860.5</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV56979927</a>
ENST00000514860.5:c.1480+2T>C	<a href="#">4:165313368-165313368</a>	C	splice_donor_variant	KLHL2	<a href="#">ENSG00000109466</a>	Transcript	<a href="#">ENST00000538127.5</a>	protein_coding	-	-	-	-	-	-	<a href="#">COSV56979927</a>

## ESEfinder

Para las predicciones dónde se encuentra la posición de interés solo se obtienen puntuaciones positivas para dos resultados y para las matrices 5' (475 y 481):

475 (-375)	GGCGGAGTGGAGCAGgtacatgtgaacctg	8.48490	475 (-375)	GGCGGAGTGGAGCAGgtacatgtgaacctg	-1.99810	475 (-375)	GGCGGAGTGGAGCAGgtacatgtgaacctg	8.32960	475 (-375)	GGCGGAGTGGAGCAGgtacatgtgaacctg	-2.27910
481 (-369)	GTGGAGCAGgtacatgtgaacctgttttag	4.63170	481 (-369)	GTGGAGCAGgtacatgtgaacctgttttag	-13.37900	481 (-369)	GTGGAGCAGgtacatgtgaacctgttttag	4.43590	481 (-369)	GTGGAGCAGgtacatgtgaacctgttttag	-14.25140

Si observamos los resultados equivalentes para la secuencia mutante, se observa que en la predicción 475 bajan considerablemente las puntuaciones, mientras que la 481, las puntuaciones aumentan ligeramente.

475 (-375)	GGCGGAGTGGAGCAGgcacatgtgaacctg	1.54880	475 (-375)	GGCGGAGTGGAGCAGgcacatgtgaacctg	-2.46150	475 (-375)	GGCGGAGTGGAGCAGgcacatgtgaacctg	1.21950	475 (-375)	GGCGGAGTGGAGCAGgcacatgtgaacctg	-2.74460
481 (-369)	GTGGAGCAGgcacatgtgaacctgttttag	5.05110	481 (-369)	GTGGAGCAGgcacatgtgaacctgttttag	-13.80800	481 (-369)	GTGGAGCAGgcacatgtgaacctgttttag	4.76680	481 (-369)	GTGGAGCAGgcacatgtgaacctgttttag	-14.64560

Por lo tanto, lo más probable es que se esté debilitando un sitio *donor* (475), lo que puede afectar al *splicing*.