

Ejemplo comparación de resultados predictores in silico

Cambio de estudio STK11 c.465-1G>T (chr19:1220372 G/T, COSV58820358 o NM_000455.5: c.465-1G>T)

Exón 5 e intrones adyacentes:

```
gggaggggctgtgtgtccagcaagactttgggggtgcagccggcctgtggcccaaggaa  
aatgagacctgtggacatccggggccctgccagacgtggctcgggacggagggtgcc  
actgcaggcgcaggtgtgggtccctgtctggacctagcctttcctcgtcctgtgtgcctg  
gacttctgtgacttcccagctgggcctgtggtgttttgggaggctcccaggcagctgcaaa  
ggggaccctgtgagggggcagggaggcctcgggccaggacgggtgtgtgctgcccgcag  
GTACTTCTGTCAGCTGATTGACGGCCTGGAGTACCTGCATAGCCAGGCATTGTGCACAA  
GGACATCAAGCCGGGAACCTGCTGCTCACACCGGTGGCACCTCAAAATCTCCGACCT  
GGGCGTGGCCGAG  
gtaggcacgtgctaggggggggcctgggggcgccctcccgggcactccetgagggctgc  
acggcacggccacag
```

El cambio se encuentra en la posición justo antes del exón 5 (la **g** en color rojo).

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'
134		2	+	0.61	GCAGGCGCAG	^GTGTGGCTCC		
434		0	+	0.00	CGTGGCCGAG	^GTAGGCACGT		

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
334		175		0	-	0.55	GGCTATGCAG	^GTACTCCAGG		
304		205		0	-	0.70	GCTGACAGAA	^GTACCTGCGG		

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
255		2	+	0.07	CCCCTGTGAG	^GGGCAGGGAG		
261		2	+	0.18	TGAGGGGCAG	^GGAGGCCTCG		
265		0	+	0.19	GGGCAGGGAG	^GCCCTCGGCC		
278		1	+	0.34	TCGGCCCCAG	^GACGGGTGTG		
300		2	+	0.83	CTGCCCCGAG	^GTACTTCTGT		
313		0	+	0.18	CTTCTGTCAG	^CTGATTGACG		

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
412		97		1	-	0.19	CAGGTCGGAG	^ATTTTGAGGG		
404		105		0	-	0.18	AGATTTTGAG	^GGTGCCACCG		
249		260		0	-	0.50	CCCCTCACAG	^GGGTCCCCTT		
100		409		0	-	0.07	TCCGGCCGAG	^CCACGTCTGG		

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
134		2	+	0.61	GCAGGCGCAG	^GTGTGGCTCC		
434		0	+	0.00	CGTGGCCGAG	^GTAGGCACGT		

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
334		175		0	-	0.55	GGCTATGCAG	^GTACTCCAGG		
304		205		0	-	0.70	GCTGACAGAA	^GTACATGCGG		

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
255		2	+	0.17	CCCCTGTGAG	^GGGCAGGGAG		
261		2	+	0.19	TGAGGGGCAG	^GGAGGCCTCG		
265		0	+	0.19	GGGCAGGGAG	^GCCCTCGGCC		
278		1	+	0.43	TCGGCCCCAG	^GACGGGTGTG		
313		0	+	0.18	CTTCTGTCAG	^CTGATTGACG		

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
412		97		1	-	0.19	CAGGTCGGAG	^ATTTTGAGGG		
404		105		0	-	0.18	AGATTTTGAG	^GGTGCCACCG		
249		260		0	-	0.43	CCCCTCACAG	^GGGTCCCCTT		
100		409		0	-	0.07	TCCGGCCGAG	^CCACGTCTGG		

Se pierde uno de los sitios *acceptor* (en azul) de la secuencia *wild type* en la secuencia mutante. Este coincide con el sitio *acceptor* del exón, por lo tanto, se van a producir cambios en el *splicing*. En el caso de que se empleará cualquiera de los otros sitios predichos (y siempre que se emplee el mismo *donor* del *splicing* normal) se generaría (por orden de aparición) la inclusión de 45, 39, 35 o 22 nucleótidos en el exón o se perderían los primeros 13 nucleótidos del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
294	308	0.94	cccgag	gtacttct
427	441	0.96	ggcgag	gtaggcac

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
179	219	0.56	tggacttctgtgacttccc	agctgggcctgtggtgtttggg

Donor site predictions for mut :

Start	End	Score	Exon	Intron
427	441	0.96	ggcgag	gtaggcac

Acceptor site predictions for mut :



Start	End	Score	Intron	Exon
179	219	0.56	tggacttctgtgacttccc	agctgggcctgtggtgtttggg

Se pierde uno de los sitios *donor* (en azul) de la secuencia *wild type* en la secuencia mutante. Este no coincide con el sitio *donor* del exón, por lo tanto, no se va a tener en cuenta.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ccgca(g/t)gtact	ggtact	tgtact	26576	52%

Human Splicing Finder

 Broken WT Acceptor Site		Alteration of the WT Acceptor site, most probably affecting splicing	
Algorithm/Matix	position	sequences	variation
MaxEnt Acceptor site	chr19:1220353	- REF : CGGGTGTGTGCTGCCCCGCA G GTA - ALT : CGGGTGTGTGCTGCCCCGCA T GTA	10.66 > 2.06 => -80.68%
HSF Acceptor site (matrix AG)	chr19:1220361	- REF : TGCTGCCCCGCA G GT - ALT : TGCTGCCCCGCA T GT	84.34 > 56.47 => -33.04%
 New Acceptor splice site		Activation of a cryptic Acceptor site. Potential alteration of splicing	
Algorithm/Matix	position	sequences	variation
MaxEnt Acceptor site	chr19:1220366	- REF : CCCGCA G GTA C TTCTGTCAGCTG - ALT : CCCGCA T GTA C TTCTGTCAGCTG	2.1 > 5.13 => 144.29%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	12	32	ctgtcagct		-1.10006779203	0.481481481481	27	0	0	0	-1.5825009
wt	12	27	agctgattg		1.05406873574	0.5	22	0	0	0	-0.41658214
wt	12	23	gattgacgg		1.94478380002	0.5	18	0	0	0	0.18536822
mut	12	32	ctgtcagct		-1.10006779203	0.481481481481	27	0	0	0	-1.5825009
mut	12	27	agctgattg		1.05406873574	0.5	22	0	0	0	-0.41658214
mut	12	23	gattgacgg		1.94478380002	0.5	18	0	0	0	0.18536822

Variant Effect Predictor tool

ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	splice_acceptor_variant	STK11	ENSG00000118046	Transcript	ENST00000326873.12	protein_coding	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	upstream_gene_variant	STK11	ENSG00000118046	Transcript	ENST00000585465.2	protein_coding	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	downstream_gene_variant	STK11	ENSG00000118046	Transcript	ENST00000585748.2	protein_coding	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	splice_acceptor_variant	STK11	ENSG00000118046	Transcript	ENST00000585851.1	protein_coding	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	splice_acceptor_variant	STK11	ENSG00000118046	Transcript	ENST00000586243.5	protein_coding	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	splice_acceptor_variant, non_coding_transcript_variant	STK11	ENSG00000118046	Transcript	ENST00000586358.5	retained_intron	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	splice_acceptor_variant, non_coding_transcript_variant	STK11	ENSG00000118046	Transcript	ENST00000589152.5	retained_intron	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	non_coding_transcript_exon_variant	STK11	ENSG00000118046	Transcript	ENST00000591133.2	retained_intron	1/4	360	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	downstream_gene_variant	STK11	ENSG00000118046	Transcript	ENST00000593219.5	nonsense_mediated_decay	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934
ENST00000586243.5:c.465-1G>T	19:1220372-1220372	T	splice_acceptor_variant	STK11	ENSG00000118046	Transcript	ENST00000652231.1	protein_coding	-	-	-	-	-	-	CS982373, CX075745, COSV58820358, COSV58822934

ESEfinder

Se observan tres predicciones con puntuaciones positivas: una solo para las matrices 5' (271), una solo para las matrices 3' (299) y una con puntuaciones positivas para todas las matrices (286):

