

Ejemplo comparación de resultados predictores in silico

Cambio de estudio ARHGAP28 c.955-1G>T (chr18:6873408 G/T, COSV51634121 o NM_001366230.1: c.955-1G>T)

Exón 8 e intrones adyacentes:

```
tgacatgtatgttgtttgacaatatgggtgttccttataaagaaggcagtttccacagt
tgcaatggcaatgttatcataatacttttagactgctgttttgcttctaattagcatagcag
tgatctgcacattgttctaaccacatattcataatttctaggatgcagcatgctccttag
agatgttttccaaagaacacgttctactgcataagatttagagtgtccaggtggcatttga
gtgaacgcataataattttccctttgccataaaaaataaataccttcactgtgtgttttag
AATTTAATGTTTCAGAAAAACAGATTGGCTTAACTGAAGCAGGACATCTCTGCTGAA
GACATGAAGAAAATCCGCATCTCTCTCTGATTGAATTGACTGCCTTTTTTGATGCCTTT
GGAATTCAACTGAAAAGCAACAAAACAGAGAAAGTAAAAGGACGAG
gtaac taaga agactgattgcttctggctttaacactttgacacagtggaataagataat
gcttttctaattctttttttttttccccctttactgtctcacaatgaag
```

El cambio se encuentra en la primera posición justo antes del exón 8 (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'
454		0	+	0.71	AACAGAGAAA	^	GTAAAAGGAC	
467		1	+	0.93	AAAGGACGAG	^	GTAACCTAAGA	H

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
282		294		0	-	0.34	CACAGTGAAG	^	GTATTTATTT	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
161		0	+	0.16	ATATTTCTAG	^	GATGCAGCAT	
300		0	+	0.87	TGTGTTTTAG	^	AAATTTAATG	
315		0	+	0.20	TAATGTTTCTAG	^	AAAACCCAGAT	
323		2	+	0.31	AGAAAACCCAG	^	ATTTGGCTTA	
340		1	+	0.19	TAACTGAAG	^	CAGGAGATCT	
343		1	+	0.19	ACTGAAGCAG	^	GAGATCTGTC	
346		1	+	0.19	GAAGCAGGAG	^	ATCTGTCTGC	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
430		146		0	-	0.97	TCCTTTTCTAG	^	TTGAATTCCA	
356		220		2	-	0.16	ATGTCTTCTAG	^	CAGACAGATC	

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
454		0	+	0.71	AACAGAGAAA	^	GTAAAAGGAC	
467		1	+	0.93	AAAGGACGAG	^	GTAACCTAAGA	H

Donor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
282		294		0	-	0.34	CACAGTGAAG	^	GTATTTATTT	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
161		0	+	0.16	ATATTTCTAG	^	GATGCAGCAT	
315		0	+	0.31	TAATGTTCTAG	^	AAAACCCAGAT	
323		2	+	0.34	AGAAAACCCAG	^	ATTTGGCTTA	
340		1	+	0.19	TAACTGAAG	^	CAGGAGATCT	
343		1	+	0.19	ACTGAAGCAG	^	GAGATCTGTC	
346		1	+	0.18	GAAGCAGGAG	^	ATCTGTCTGC	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
430		146		0	-	0.97	TCCTTTTCTAG	^	TTGAATTCCA	

Se pierde un sitio *acceptor* (en rojo) en la secuencia mutante. Este es el sitio *acceptor* del exón, por lo que, si no se emplea este sitio, habrá cambios en el *splicing*. Dependiendo cuál de los otros se use (por orden de aparición y suponiendo que el sitio *donor* empleado sigue siendo el mismo), se incluirán 139 pb al exón o se perderán 15, 23, 40, 43 o 46 pb del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.2.148.425270.0 :

Start	End	Score	Exon	Intron
460	474	0.99	ggacgag	gt ^{aactaa}

Donor site predictions for 10.42.3.123.425282.0 :

Start	End	Score	Exon	Intron
460	474	0.99	ggacgag	gt ^{aactaa}

Acceptor site predictions for 10.42.2.148.425270.0 :

Start	End	Score	Intron	Exon
141	181	0.46	accacatattcatatttct	aggatgcagcatgctccttaga
194	234	0.70	agaacacgttctactgcat	agatttagagtgtccaggtggc
280	320	0.51	taccttcactgtgtgtttt	agaaatttaatgttcagaaaac

Acceptor site predictions for 10.42.3.123.425282.0 :


Start	End	Score	Intron	Exon
141	181	0.46	accacatattcatatttct	aggatgcagcatgctccttaga
194	234	0.70	agaacacgttctactgcat	agatttagagtgtccaggtggc


Se pierde uno de los sitios *acceptor* (en rojo). Este es el sitio *acceptor* del exón, por lo que, si no se emplea este sitio, habrá cambios en el *splicing*.
Dependiendo cuál de los otros se use (por orden de aparición y suponiendo que el sitio *donor* empleado sigue siendo el mismo), se incluirán 139 o 86 pb al exón.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
tttta(g't)aaatt	ttttag	ttttat	39392	90%

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	chr18:6873389	- REF : ACCTTCACTGTGTGTTTAA G AAA - ALT : ACCTTCACTGTGTGTTTAA T AAA	9.83 > 1.24 => -87.39%
HSF Acceptor site (matrix AG)	chr18:6873397	- REF : TGTGTGTTTAA G AA - ALT : TGTGTGTTTAA T AA	81.89 > 54.02 => -34.03%

 New Acceptor splice site	Activation of a cryptic Acceptor site. Potential alteration of splicing		
Algorithm/Matix	position	sequences	variation
MaxEnt Acceptor site	chr18:6873404	- REF : TT T AGAAATTTAATGTTTCAGAAA - ALT : TTTA T AAATTTAATGTTTCAGAAA	3.09 > 5.46 => 76.7%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	28	65	ccttcactg		0.714391224453		0.433333333333	1	11	17	0.91652082
wt	28	53	gttttagaa		-5.24970506948		0.375	48	0		-4.5709351
wt	28	45	aatttaatg		-3.6330989497		0.375	40	0		-3.4315709
wt	28	44	atttaatgt		-0.105843574772		0.384615384615	39	0	0	-1.9840799
wt	28	38	tgttcagaa		-1.79782458978		0.333333333333	33	0	0	-2.2833431
wt	28	19	ggcttaact		-1.04951052827		0.214285714286	14	0	0	-0.82612356
wt	28	18	gcttaactg		1.22807147245		0.230769230769	13	0	0	0.13427943
mut	12	66	ccttcactg		0.714391224453		0.459016393443	1	13	18	0.93413153
mut	12	54	gttttataa		-4.52202237074		0.408163265306	49	0	0	-4.3386005
mut	12	51	ttataaatt		-1.65516640541		0.413043478261	46	0	0	-3.0246198
mut	12	46	aatttaatg		-3.6330989497		0.390243902439	41	0	0	-3.4899459
mut	12	45	atttaatgt		-0.105843574772		0.4	40	0		-2.0424095
mut	12	39	tgttcagaa		-1.79782458978		0.352941176471	34	0	0	-2.3403087
mut	12	20	ggcttaact		-1.04951052827		0.266666666667	15	0	0	-0.87250437
mut	12	19	gcttaactg		1.22807147245		0.285714285714	14	0	0	0.08872675
mut	12	15	aactgaagc		0.581778608334		0.2	10	0		0.061182631

Las diferencias entre ambas es que, a raíz de la mutación, uno de los BP cambia solo en la posición donde se encuentra el cambio y aparece un nuevo BP (ttataaatt). Como sigue teniendo puntuación positiva, no se va a tener en cuenta.

Variant Effect Predictor tool

ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	splice_acceptor_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000262227.7	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	splice_acceptor_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000314319.7	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	splice_acceptor_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000383472.9	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	splice_acceptor_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000419673.6	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	splice_acceptor_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000531294.5	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	splice_acceptor_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000532996.5	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	downstream_gene_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000577524.5	nonsense_mediated_decay	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	upstream_gene_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000579245.1	retained_intron	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	upstream_gene_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000579689.1	protein_coding	-	-	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	non_coding_transcript_exon_variant	AP005210.1	ENSG00000263797	Transcript	ENST00000583659.1	lncRNA	2/2	535	COSV51634121
ENST00000383472.8:c.955-1G>T	18:6873408-6873408	T	downstream_gene_variant	ARHGAP28	ENSG00000088756	Transcript	ENST00000584287.5	nonsense_mediated_decay	-	-	COSV51634121

ESEfinder

En las predicciones para la secuencia *wild type*, solo se da una predicción con puntuación positiva para las matrices 5' (275) y una predicción con puntuaciones positivas para las matrices 3' (286):

275 (-301)	ataaataccttcactgtgtgttttagAAAT	1.18950	275 (-301)	ataaataccttcactgtgtgttttagAAAT	-29.11580	275 (-301)	ataaataccttcactgtgtgttttagAAAT	0.77090	275 (-301)	ataaataccttcactgtgtgttttagAAAT	-28.52820
286 (-290)	cactgtgtgttttagAAATTTAATGTTTCAG	-23.86220	286 (-290)	cactgtgtgttttagAAATTTAATGTTTCAG	3.35920	286 (-290)	cactgtgtgttttagAAATTTAATGTTTCAG	-20.56570	286 (-290)	cactgtgtgttttagAAATTTAATGTTTCAG	3.19960

Las puntuaciones para 275 en la secuencia mutante son muy parecidas a las anteriores, mientras que para 286 las puntuaciones en las matrices 3' son negativas, por lo tanto lo más probable es que se esté perdiendo un sitio *acceptor* en la secuencia mutante:

275 (-301)	ataaatacettcactgtgtgttttatAAAT	1.02380	275 (-301)	ataaatacettcactgtgtgttttatAAAT	-29.56950	275 (-301)	ataaatacettcactgtgtgttttatAAAT	0.39890	275 (-301)	ataaatacettcactgtgtgttttatAAAT	-28.92470
286 (-290)	cactgtgtgttttatAAATTTAATGTTTCAG	-27.83200	286 (-290)	cactgtgtgttttatAAATTTAATGTTTCAG	-8.72840	286 (-290)	cactgtgtgttttatAAATTTAATGTTTCAG	-24.48890	286 (-290)	cactgtgtgttttatAAATTTAATGTTTCAG	-9.75380