

Ejemplo comparación de resultados predictores in silico

Cambio de estudio KRT17 c.433-195T>C (chr17:41623227 T/C, COSV60860377 o NM_000422.3: c.433-195T>C)

Exón 2 e intrones adyacentes:

```
ctgccctggggacatttttccaattcttttctggttgctcatactccaggccagctgtct
ctttctcctttaaggcgagcctgccatgggggtctggtggggtacgagtatcgggggaa
gaagaggcacctttcaggccttcagacctctgtttgcccctctctgccaataatacagc
acgggggcaagggagggggctgggcgagaagagaggccctgaggcaggaagatctgctcaga
accctggtgtggggctcaggccaccccatccaatgactgactactctccttcctctcag
ATCCTCACAGCCACCGTGGACAATGCCAACCATCCTGCTACAGATTGACCAATGCCCGTCTG
GCTGCTGATGACTTCCGCACCAA
gtgagtcctagctgtggccttgggcagcctgggccagctggcgaggatctcagggtacc
cctctgacccccaggattccttggtgcttgtggcaaggcccagagctcaggggtggggc
agtcttaggagccccacctccttagtccaggatgcagtgaaggcagccagttctgaaggt
gctgagcttaggcagggaatagaaagagaaggaggggagggcgggagggcgggagggcagagag
aagtaaggaagctggtggggcgtgggatctggccctgtgatggttcccgggtccggggt
ggaattcgtttcaactcgacccctctcatcagcccttccaactccttagagtccctggcaaa
atgaaggcaggtgagcagccaggacctgatctgcaggtccaagagcctgggctgaagt
cttgattcccacggcag
```

El cambio se encuentra en la segunda línea del intrón 1 (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'
384	2	+	0.86	TCCGCACCAA	^	GTGAGTCCTA	
439	0	+	0.70	GGATCTCAGG	^	GTACCCCTCC	
754	2	+	0.00	ATGAAGGCAG	^	GTGAGCAGCC	

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
314	508	1	-	0.31	ATTGTCCACG	^	GTGGCTGTGA	

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
73	0	+	0.00	CTCCTTTAAG	^	GCCGAGCCTG	
258	0	+	0.17	GTGGGCTCAG	^	CCACCCCAT	
300	0	+	1.00	TTCTCCTCAG	^	ATCCTCACAG H	
310	1	+	0.77	ATCCTCACAG	^	CCACCGTGGA	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
736	86	0	-	0.00	ATTTTGCCAG	^	GACTCTAAGG	
421	401	0	-	0.17	CCTCCGCCAG	^	CTGGCCAGG	
217	605	0	-	0.15	CCTGCCTCAG	^	GGCCTCTCTT	
105	717	1	-	0.25	CGATACTCAG	^	TACCCACCA	

Donor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
384	2	+	0.86	TCCGCACCAA	^	GTGAGTCCTA	
439	0	+	0.70	GGATCTCAGG	^	GTACCCCTCC	
754	2	+	0.00	ATGAAGGCAG	^	GTGAGCAGCC	

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
314	508	1	-	0.31	ATTGTCCACG	^	GTGGCTGTGA	

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
73	0	+	0.00	CTCCTTTAAG	^	GCCGAGCCTG	
258	0	+	0.17	GTGGGCTCAG	^	CCACCCCAT	
300	0	+	1.00	TTCTCCTCAG	^	ATCCTCACAG H	
310	1	+	0.77	ATCCTCACAG	^	CCACCGTGGA	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
736	86	0	-	0.00	ATTTTGCCAG	^	GACTCTAAGG	
421	401	0	-	0.17	CCTCCGCCAG	^	CTGGCCAGG	
217	605	0	-	0.15	CCTGCCTCAG	^	GGCCTCTCTT	

Desaparece un sitio *acceptor* en la secuencia mutante. Este se encuentra en medio del intrón coincidiendo con la región donde se está la posición de estudio. Como no participa en el *splicing* normal, es probable que no afecte a cuando existe la mutación.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.0.139.574686.0 :

Start	End	Score	Exon	Intron
377	391	0.99	gcaccaa	gt gagtcc
619	633	0.83	gagagaa	gt aaggaa
747	761	0.98	aaggcag	gt gagcag

Acceptor site predictions for 10.42.0.139.574686.0 :

Start	End	Score	Intron	Exon
53	93	0.99	agctgtctcttctccttta	ag gccgagcctgccatgggggt
125	165	0.69	aggcacctttcagcccttc	ag actcctgtttgccctcctc
280	320	0.99	actactctcccttctcctc	ag atcctcacagccaccgtgga
290	330	0.70	cttctctcagatcctcac	ag ccaccgtggacaatgccaac
437	477	0.54	gggtacccctcctgacccc	ag gattccttggttgcttggtg
512	552	0.44	gagccccactccttagtcc	ag gatgcagtgaaggcagccag
711	751	0.88	tcagcccttccaactcctt	ag agtccttgcaaaatgaaggc

Donor site predictions for 10.42.1.119.574698.0 :

Start	End	Score	Exon	Intron
377	391	0.99	gcaccaa	gt gagtcc
619	633	0.83	gagagaa	gt aaggaa
747	761	0.98	aaggcag	gt gagcag

Acceptor site predictions for 10.42.1.119.574698.0 :

Start	End	Score	Intron	Exon
53	93	0.99	agctgtctcttctccttta	ag gccgagcctgccatgggggt
125	165	0.69	aggcacctttcagcccttc	ag actcctgtttgccctcctc
280	320	0.99	actactctcccttctcctc	ag atcctcacagccaccgtgga
290	330	0.70	cttctctcagatcctcac	ag ccaccgtggacaatgccaac
437	477	0.54	gggtacccctcctgacccc	ag gattccttggttgcttggtg
512	552	0.44	gagccccactccttagtcc	ag gatgcagtgaaggcagccag
711	751	0.88	tcagcccttccaactcctt	ag agtccttgcaaaatgaaggc

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ggtac(t/c)gagta	ctgagt	ccgagt	29948	74%

Human Splicing Finder

Alteration of auxiliary sequences		Significant alteration of ESE / ESS motifs ratio (4)
Algorithm/Matix	position	sequence
IIE (ESS Site Broken)	chr17:41623228	CTGAGT
EIE (New ESE Site)	chr17:41623229	ACCGAG
EIE (New ESE Site)	chr17:41623230	TACCGA
ESE_SC35 (New ESE Site)	chr17:41623231	GTACCGAG

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	38	143	gcctcatcac		0.955449034289	0.557971014493	14	16	37 0.41459937
wt	38	113	cctttaagg		-3.8132972486	0.5	57	16	25 -4.3049311
wt	38	112	ctttaaggc		-1.44345925186	0.504672897196	56	16	25 -3.31222
wt	38	76	tactgagta		0.213554412294	0.549295774648	20	16	25 -0.37026935
wt	38	48	ctttcagcc		-0.799167552003	0.674418604651	11	23	42 0.0016673865
wt	38	40	ccttcagac		-1.03302437877	0.657142857143	3	23	42 0.41090838
mut	38	143	gcctcatcac		0.955449034289	0.557971014493	14	16	37 0.41459937
mut	38	113	cctttaagg		-3.8132972486	0.5	57	16	25 -4.3049311
mut	38	112	ctttaaggc		-1.44345925186	0.504672897196	56	16	25 -3.31222
mut	38	48	ctttcagcc		-0.799167552003	0.674418604651	11	23	42 0.0016673865
mut	38	40	ccttcagac		-1.03302437877	0.657142857143	3	23	42 0.41090838

Desaparece un BP en la secuencia mutante, pero, como tiene puntuación negativa, no se va a tener en cuenta.

Variant Effect Predictor tool

ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	intron_variant	KRT17	ENSG00000128422	Transcript	ENST00000311208.13	protein_coding	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	downstream_gene_variant	KRT42P	ENSG00000214514	Transcript	ENST00000398469.5	processed_transcript	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	downstream_gene_variant	KRT42P	ENSG00000214514	Transcript	ENST00000438131.5	processed_transcript	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	downstream_gene_variant	KRT42P	ENSG00000214514	Transcript	ENST00000458343.2	transcribed_unprocessed_pseudogene	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	intron_variant	KRT17	ENSG00000128422	Transcript	ENST00000463128.5	protein_coding	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	intron_variant, non_coding_transcript_variant	KRT17	ENSG00000128422	Transcript	ENST00000491673.1	retained_intron	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	intron_variant, non_coding_transcript_variant	KRT17	ENSG00000128422	Transcript	ENST00000493253.5	retained_intron	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	intron_variant	KRT17	ENSG00000128422	Transcript	ENST00000540235.5	protein_coding	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	intron_variant	KRT17	ENSG00000128422	Transcript	ENST00000577817.3	protein_coding	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	downstream_gene_variant	KRT42P	ENSG00000214514	Transcript	ENST00000587335.1	retained_intron	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	non_coding_transcript_exon_variant	KRT17	ENSG00000128422	Transcript	ENST00000590038.1	retained_intron	1/3	42	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	upstream_gene_variant	KRT17	ENSG00000128422	Transcript	ENST00000648859.1	protein_coding	-	-	COSV60860377
ENST00000311208.12:c.433-195T>C	17:41623227-41623227	G	upstream_gene_variant	KRT17	ENSG00000128422	Transcript	ENST00000649249.1	retained_intron	-	-	COSV60860377

ESEfinder

Se obtienen 3 predicciones para la secuencia WT con puntuaciones positivas en las matrices 5'SS (77, 82 y 87).

77 (-745)	gagcctgccatgggggtctggtgggtact	2.86700	77 (-745)	gagcctgccatgggggtctggtgggtact	-19.31160	77 (-745)	gagcctgccatgggggtctggtgggtact	2.63300	77 (-745)	gagcctgccatgggggtctggtgggtact	-17.83390
82 (-740)	tgccatgggggtctggtgggtactgagta	6.07880	82 (-740)	tgccatgggggtctggtgggtactgagta	-12.03180	82 (-740)	tgccatgggggtctggtgggtactgagta	5.65070	82 (-740)	tgccatgggggtctggtgggtactgagta	-13.20530
87 (-735)	tgggggtctggtgggtactgagtatcggg	1.78880	87 (-735)	tgggggtctggtgggtactgagtatcggg	-22.34290	87 (-735)	tgggggtctggtgggtactgagtatcggg	1.43500	87 (-735)	tgggggtctggtgggtactgagtatcggg	-20.97460

Cuando comprobamos las predicciones equivalentes para la secuencia mutante, se observa que las puntuaciones se alteran ligeramente, pero no lo suficiente para tenerse en cuenta.

77 (-745)	gagcctgccatgggggtctggtgggtacc	2.99170	77 (-745)	gagcctgccatgggggtctggtgggtacc	-18.71070	77 (-745)	gagcctgccatgggggtctggtgggtacc	2.90370	77 (-745)	gagcctgccatgggggtctggtgggtacc	-17.24630
82 (-740)	tgccatgggggtctggtgggtaccgagta	6.27850	82 (-740)	tgccatgggggtctggtgggtaccgagta	-11.43510	82 (-740)	tgccatgggggtctggtgggtaccgagta	5.98050	82 (-740)	tgccatgggggtctggtgggtaccgagta	-12.61910
87 (-735)	tgggggtctggtgggtaccgagtatcggg	1.73470	87 (-735)	tgggggtctggtgggtaccgagtatcggg	-21.97720	87 (-735)	tgggggtctggtgggtaccgagtatcggg	1.60220	87 (-735)	tgggggtctggtgggtaccgagtatcggg	-20.64260

Por otro lado, en el caso de la predicción 94 ocurre que en la secuencia mutante tiene puntuación positiva y en la WT puntuación negativa. Puede, por tanto, que se esté activando un sitio *donor* (aunque es poco probable dada la puntuación tan baja).

94 (-728)	ctggtgggtactgagtatcgggggaagaa	-1.90020	94 (-728)	ctggtgggtactgagtatcgggggaagaa	-27.87910	94 (-728)	ctggtgggtactgagtatcgggggaagaa	-1.96270	94 (-728)	ctggtgggtactgagtatcgggggaagaa	-29.57680
94 (-728)	ctggtgggtaccgagtatcgggggaagaa	0.25060	94 (-728)	ctggtgggtaccgagtatcgggggaagaa	-26.29080	94 (-728)	ctggtgggtaccgagtatcgggggaagaa	0.13920	94 (-728)	ctggtgggtaccgagtatcgggggaagaa	-27.96310