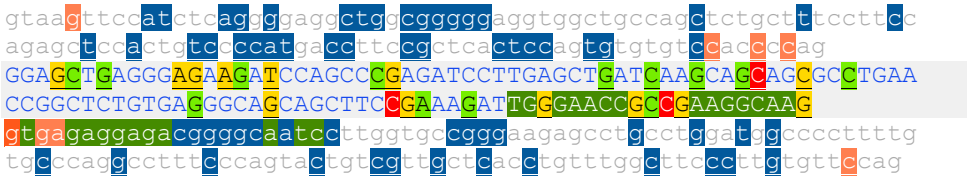
Ejemplo comparación de resultados predictores in sillico

Cambio de estudio ELMO2 c.1684C>T (chr20:46371588 C/T, COSV51682670 o NM_133171.5:c.1684C>T)

Exón 18 e intrones adyacentes:



El cambio se encuentra en la segunda línea del exón 7 (la segunda **c** 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'
223	2	+	0.70	ACCGCCGAAG^GCAAGGTGAG
228	1	+	1.00	CGAAGGCAAG^GTGAGAGGAG H

Donor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' exon intron 3'
321	26	-	-	0.00	AGCCAAACAG^GTGAGCAACG

Acceptor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' intron exon 3'
62	0	+	0.00	TTCCTTCCAG^AGCTCCACTG
98	1	+	0.96	CTCACTCCAG^TGTGTGTCCA H
114	2	+	1.00	TCCACCCCAG^GGAGCTGAGG H
118	0	+	0.34	CCCCAGGGAG^CTGAGGGAGA
123	2	+	0.20	GGGAGCTGAG^GGAGAAGATC
130	0	+	0.19	GAGGGAGAAG^ATCCAGCCCG
136	0	+	0.17	GAAGATCCAG^CCCGAGATCC
294	0	+	0.00	TTGTGCCCAG^GCCTTTCCCA

Acceptor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' intron exon 3'
196	151	1	-	0.32	CTTTCGGAAG^CTGCTGCCCT
181	166	1	-	0.30	GCCCTCACAG^AGCCGGTTCA
179	168	0	-	0.18	CCTCACAGAG^CCGGTTCAGG

Donor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' exon intron 3'
216	1	+	0.00	ATTGGGAACC^GCTGAAGGCA
223	2	+	0.70	ACCGCTGAAG^GCAAGGTGAG
228	1	+	1.00	TGAAGGCAAG^GTGAGAGGAG H

Donor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' exon intron 3'
321	26	-	-	0.00	AGCCAAACAG^GTGAGCAACG

Acceptor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' intron exon 3'
62	0	+	0.00	TTCCTTCCAG^AGCTCCACTG
98	1	+	0.96	CTCACTCCAG^TGTGTGTCCA H
114	2	+	1.00	TCCACCCCAG^GGAGCTGAGG H
118	0	+	0.34	CCCCAGGGAG^CTGAGGGAGA
123	2	+	0.20	GGGAGCTGAG^GGAGAAGATC
130	0	+	0.19	GAGGGAGAAG^ATCCAGCCCG
136	0	+	0.07	GAAGATCCAG^CCCGAGATCC
294	0	+	0.00	TTGTGCCCAG^GCCTTTCCCA

Acceptor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' intron exon 3'
217	130	0	-	0.55	TTGCCTTCAG^CGGTTCCCAA
196	151	1	-	0.32	CTTTCGGAAG^CTGCTGCCCT
181	166	1	-	0.30	GCCCTCACAG^AGCCGGTTCA
179	168	0	-	0.18	CCTCACAGAG^CCGGTTCAGG

Aparece un nuevo sitio de *accpetor* en la secuencia mutante. Si se empleara este en vez del normal, se incluiría un exón críptico a partir del intrón 18.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.81.31.8546.0 :

Start	End	Score	Exon	Intron
221	235	0.98	aggcaag	gtgagagg

Donor site predictions for 85.53.81.31.8535.0 :

Start	End	Score	Exon	Intron
221	235	0.98	aggcaag	gtgagagg

Acceptor site predictions for 85.53.81.31.8546.0 :

Start	End	Score	Intron	Exon
42	82	0.98	ccagctctgctttccttcc	agagctccactgtccccatgac
78	118	0.73	atgaccttccgctcactcc	agtgtgtgtccaccccagggag
94	134	0.69	tccagtgtgtgtccaccccc	aggagctgagggagaagatcc
274	314	0.52	gatggccccttttgtgccc	agcctttcccagtactgtcgt
285	325	0.48	ttgtgcccaggcctttccc	agtactgtcgttgctcacctgt

Acceptor site predictions for 85.53.81.31.8535.0 :

Start	End	Score	Intron	Exon
42	82	0.98	ccagctctgctttccttcc	agagctccactgtccccatgac
78	118	0.73	atgaccttccgctcactcc	agtgtgtgtccaccccagggag
94	134	0.69	tccagtgtgtgtccaccccc	aggagctgagggagaagatcc
274	314	0.52	gatggccccttttgtgccc	agcctttcccagtactgtcgt
285	325	0.48	ttgtgcccaggcctttccc	agtactgtcgttgctcacctgt

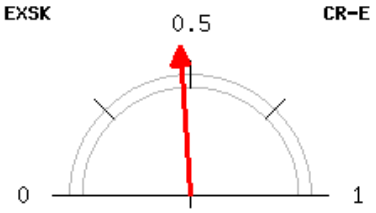
Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
accgc(c/t)gaagg	ccgaag	ctgaag	30176	75%

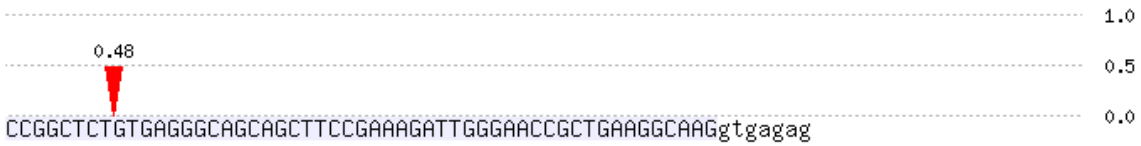
CRYP-SKIP

Results for sequence wt

Exon length (bp)	113
PESS (<=-2.62) density	0.00
NN 5'ss score density	0.09
SF2/ASF score density	17.60
FAS-ESS (hex2) density	0.00
EIE score density	536.44
Probability of cryptic splice site activation (P_{CR-E})	0.48




```
>wt
ccagtgtgtgtccaccccagGGAGCTGAGGGAGAAGATCCAGCCCGAGATCCTTGAGCTGATCAAGCAGCAGCGCCTGAA
```



Parece que hay un sitio críptico de *splicing* dentro del propio exón, pero el cambio de interés no lo toma en consideración, por lo que no debe considerar que tenga algún efecto en el *splicing*.

Human Splicing Finder

New Acceptor splice site	Activation of a cryptic Acceptor site. Potential alteration of splicing (cryptic exon activation)

Algorithm/Matix	position	sequences	variation
MaxEnt Acceptor site	chr20:46371570	- REF : TCCTCTCACCTTGCCTTCGGCGG - ALT : TCCTCTCACCTTGCCTTCAGCGG	0.9 > 8.86 => 884.44%
HSF Acceptor site (matrix AG)	chr20:46371578	- REF : CCTTGCCTTCGGCG - ALT : CCTTGCCTTCAGCG	56.96 > 84.83 => 48.93%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	41	332	atctcaggg	-0.916442994094	0.507645259939	31	15	33	-1.4479234		
wt	41	266	ccatgacct	1.9952750743	0.505747126437	1	15	26	1.5252775		
wt	41	255	cgctcactc	2.9487092245	0.492	7	14	16	1.4212038		
wt	41	225	agctgaggg	-0.460375944066	0.486363636364	157	14	25	-9.3263402		
wt	41	197	ccttgagct	-0.0812429870699	0.5	129	14	25	-7.4011346		
wt	41	192	agctgatca	0.985441492543	0.502673796791	124	14	25	-6.666122		
wt	41	189	tgatcaagc	-2.40363784403	0.5	121	14	25	-7.8040754		
wt	41	174	gcctgaacc	1.574989329	0.508875739645	106	14	25	-5.2939134		
wt	41	161	ctgtgaggg	-1.31107154938	0.50641025641	93	14	25	-5.6018607		
wt	41	116	aggtgagag	-1.62636937037	0.576576576577	48	14	25	-2.8542286		
wt	41	27	tgctcacct	2.8622433933	0.772727272727	1	22	35	2.0348059		

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
mut	41	332	atctcaggg	-0.916442994094	0.507645259939	31	15	33	-1.4479234		
mut	41	266	ccatgacct	1.9952750743	0.505747126437	1	15	26	1.5252775		
mut	41	255	cgctcactc	2.9487092245	0.492	7	14	16	1.4212038		
mut	41	225	agctgaggg	-0.460375944066	0.486363636364	157	14	25	-9.3263402		
mut	41	197	ccttgagct	-0.0812429870699	0.5	129	14	25	-7.4011346		
mut	41	192	agctgatca	0.985441492543	0.502673796791	124	14	25	-6.666122		
mut	41	189	tgatcaagc	-2.40363784403	0.5	121	14	25	-7.8040754		
mut	41	174	gcctgaacc	1.574989329	0.508875739645	106	14	25	-5.2939134		
mut	41	161	ctgtgaggg	-1.31107154938	0.50641025641	93	14	25	-5.6018607		
mut	41	127	cgctgaagg	1.47027699801	0.540983606557	59	14	25	-2.3495222		
mut	41	116	aggtgagag	-1.62636937037	0.576576576577	48	14	25	-2.8542286		
mut	41	27	tgctcacct	2.8622433933	0.772727272727	1	22	35	2.0348059		

Variant Effect Predictor tool

ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	stop_gained	ELMO2	ENSG00000062598	Transcript	ENST00000290246.11	protein_coding	18/22	1877	1684	562	R/*	CGA/TGA	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	stop_gained	ELMO2	ENSG00000062598	Transcript	ENST00000352077.6	protein_coding	16/20	1678	1678	560	R/*	CGA/TGA	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	stop_gained	ELMO2	ENSG00000062598	Transcript	ENST00000372176.5	protein_coding	18/22	1889	1420	474	R/*	CGA/TGA	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	stop_gained	ELMO2	ENSG00000062598	Transcript	ENST00000396391.5	protein_coding	17/21	1790	1684	562	R/*	CGA/TGA	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	downstream_gene_variant	ELMO2	ENSG00000062598	Transcript	ENST00000425546.5	protein_coding	-	-	-	-	-	-	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	downstream_gene_variant	ELMO2	ENSG00000062598	Transcript	ENST00000450812.5	protein_coding	-	-	-	-	-	-	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	stop_gained	ELMO2	ENSG00000062598	Transcript	ENST00000452857.5	protein_coding	3/7	385	385	129	R/*	CGA/TGA	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	downstream_gene_variant	ELMO2	ENSG00000062598	Transcript	ENST00000462491.5	processed_transcript	-	-	-	-	-	-	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	non_coding_transcript_exon_variant	ELMO2	ENSG00000062598	Transcript	ENST00000464448.1	processed_transcript	3/5	365	-	-	-	-	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	stop_gained, NMD_transcript_variant	ELMO2	ENSG00000062598	Transcript	ENST00000467800.5	nonsense_mediated_decay	7/10	858	859	287	R/*	CGA/TGA	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	downstream_gene_variant	ELMO2	ENSG00000062598	Transcript	ENST00000481852.5	processed_transcript	-	-	-	-	-	-	rs765582782, COSV51682670, COSV51685004
ENST00000290246.10:c.1684C>T	20:46371588-46371588	A	downstream_gene_variant	ELMO2	ENSG00000062598	Transcript	ENST00000488853.5	processed_transcript	-	-	-	-	-	-	rs765582782, COSV51682670, COSV51685004

ESEfinder

Aparece un resultado positive en la secuencia wt para 5'SS:

208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	5.18300	208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	-0.79050	208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	4.93320	208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	-0.95150

Sin embargo, las puntuaciones en la secuencia mutante son prácticamente iguales, por lo que se descarta:

208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	5.18300	208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	-0.79050	208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	4.93320	208 (-139)	TGGGAACCGCCGAAGGCAAGgtgagaggag	-0.95150

En la búsqueda de ESE, se da un gran cambio en las puntuaciones para 214 y para 115, que descienden considerablemente, por lo que puede que se estén perdiendo ESE.

212 (-135)	AACCGCC	-2.29814	212 (-135)	AACCGCC	-1.42027	212 (-135)	AACCGCCG	2.79662	212 (-135)	AACCGCC	-3.09030
213 (-134)	ACCGCCG	-3.85465	213 (-134)	ACCGCCG	-2.21730	213 (-134)	ACCGCCGA	-2.80453	213 (-134)	ACCGCCG	-0.88769
214 (-133)	CCGCCGA	3.62385	214 (-133)	CCGCCGA	3.93380	214 (-133)	CCGCCGAA	-2.30883	214 (-133)	CCGCCGA	-0.75980
215 (-132)	CGCCGAA	1.96647	215 (-132)	CGCCGAA	3.30575	215 (-132)	CGCCGAAG	1.22556	215 (-132)	CGCCGAA	-2.75352
216 (-131)	GCCGAAG	-5.45050	216 (-131)	GCCGAAG	-4.02641	216 (-131)	GCCGAAGG	-2.35178	216 (-131)	GCCGAAG	-0.86764
217 (-130)	CCGAAGG	1.36689	217 (-130)	CCGAAGG	1.98992	217 (-130)	CCGAAGGC	-8.58655	217 (-130)	CCGAAGG	1.78989
218 (-129)	CGAAGGC	1.02110	218 (-129)	CGAAGGC	2.30179	218 (-129)	CGAAGGCA	-2.19300	218 (-129)	CGAAGGC	0.17994

112 (-235)	cagGGAG	-0.47435	112 (-235)	cagGGAG	0.87694	112 (-235)	cagGGAGC	-7.69807	112 (-235)	cagGGAG	-3.75452
113 (-234)	agGGAGC	-4.24849	113 (-234)	agGGAGC	-2.87634	113 (-234)	agGGAGCT	-4.59846	113 (-234)	agGGAGC	-0.72252
114 (-233)	gGGAGCT	0.78324	114 (-233)	gGGAGCT	0.33674	114 (-233)	gGGAGCTG	0.54826	114 (-233)	gGGAGCT	-6.42122
115 (-232)	GGAGCTG	-3.50084	115 (-232)	GGAGCTG	-2.66251	115 (-232)	GGAGCTGA	-1.75608	115 (-232)	GGAGCTG	-2.48533
116 (-231)	GAGCTGA	1.24828	116 (-231)	GAGCTGA	-0.09615	116 (-231)	GAGCTGAG	1.08557	116 (-231)	GAGCTGA	-3.68291
117 (-230)	AGCTGAG	-2.72282	117 (-230)	AGCTGAG	-1.52012	117 (-230)	AGCTGAGG	-0.02573	117 (-230)	AGCTGAG	-3.44934
118 (-229)	GCTGAGG	-5.18278	118 (-229)	GCTGAGG	-4.31562	118 (-229)	GCTGAGGG	-2.55922	118 (-229)	GCTGAGG	1.04847

EX-SKIP

Seq	PESS (count)	FAS-ESS hex2 (count)	FAS-ESS hex3 (count)	IIE (count)	IIE (sum)	NI-ESS trusted (count)	NI-ESS all (sum)	PESE (count)	RESCUE -ESE (count)	EIE (count)	EIE (sum)	NI-ESE trusted (count)	NI-ESE all (sum)	ESS (total)	ESE (total)	ESS/ESE (ratio)
wt	0	0	0	17	217.1112	0	-1.7113	11	11	49	578.7740	67	76.8733	17	138	0.12
mut	0	0	0	17	217.1112	0	-1.7113	13	13	52	599.5437	68	78.2483	17	146	0.12

Both alleles have a comparable chance of exon skipping.

HOT-SKIP

>wt
gtaagttccatctcaggggaggctggcgggggaggtggctgccagctctgctttccttccagagctccactgtccccatg
accttccgctcactccagtgtgtgtccaccccagGGAGCTGAGGGAGAAGATCCAGCCCGAGATCCTTGAGCTGATCAAG
CAGCAGCGCCTGAACCGGCTCTGTGAGGGCAGCAGCTTCCGAAAGATTGGGAACCGCCGAAGGCAAGgtgagaggagacg
gggcaatccttggtgccgggaagagcctgcctggatggccccttttgtgcccaggcctttcccagtactgtcgttgctca
cctgtttggcttcccttgtgttccag

Mutation(s) E+71G>T, E+70T>G and E+110C>T have the highest probability of exon skipping.