

Ejemplo comparación de resultados predictores in silico

Cambio de estudio EPAS1 c.2401C>A (chr2:46382538 C/A, rs1265031766 o NM_001430.5: c.2401C>A)

Exón 15 e intrones adyacentes:

```
tctgacttagatgatgccatcagaggggtcctgaagggttggtctgtagttctgggtgactctg
agcaccttttataaaggaaaggatggttagggcttccctggcctctccctcctcaggcc
aatgctaccgtcactctctgactttgggtcttctag
ATAAGTTCACCCAAAACCCCATGAGGGGCTGGGCCATCCCCTGAGACATCTGCCGCTGC
CAAGCCTCCATCTGCCATCAGTCCCGGGGAGAACAGCAAGAGCAGCTTCCCCACAGT
GCTACGCCACCCAGTACAGGACTACAGCCTGTCTGTAGCCACAAGGTGTCAG
gtgggtgtgcccaggatctgtcagcccatcccaggattcgatgcccagggaagccacg
tctactttttttccagcgtctgcacagtgccaggcaggggaggtgcttgacttgaagtc
accatagagggtcaggtctctctggatttagggaaactctcacaagtctcttgtag
```

El cambio se encuentra en la segunda fila del exón 15 (la c en color rojo y subrayada en amarillo).

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'
323		0	+	0.71	AGCCCACAAG	^	GTGTCAGGTG	
330		1	+	0.94	AAGGTGTCAG	^	GTGGGTGTGC	H

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'
117		2	+	0.69	CCTCCCTCAG	^	GCCAATGCTA	
220		0	+	0.25	GCTGCCACAG	^	CCTCCATCTG	
237		2	+	0.34	CTGCCATCAG	^	TCCCGGGGAG	
247		0	+	0.32	TCCCGGGGAG	^	AACAGCAAGA	
252		2	+	0.20	GGGAGAACAG	^	CAAGAGCAGG	
256		0	+	0.19	GAACAGCAAG	^	AGCAGGTTCC	
258		2	+	0.19	ACAGCAAGAG	^	CAGGTTCCCC	
261		2	+	0.18	GCAAGAGCAG	^	GTTCCCCACA	
274		0	+	0.14	CCCCCACAG	^	TGCTACGCCA	
364		0	+	0.27	CCCATCCCAG	^	GATTCGATGC	
405		0	+	0.71	TTTTTCCAG	^	CGTCTGCACA	
416		2	+	0.00	GTCTGCACAG	^	TGCCAGGCAC	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
66		444		1	-	0.00	TTTATAAAAG	^	GTGCTCAGAG	

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
323		0	+	0.71	AGCCCACAAG	^	GTGTCAGGTG	
330		1	+	0.94	AAGGTGTCAG	^	GTGGGTGTGC	H

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'
117		2	+	0.69	CCTCCCTCAG	^	GCCAATGCTA	
220		0	+	0.31	GCTGCCACAG	^	CCTCCATCTG	
237		2	+	0.42	CTGCCATCAG	^	TCCCGGGGAG	
247		0	+	0.34	TCCCGGGGAG	^	AACAGCAAGA	
252		2	+	0.20	GGGAGAACAG	^	CAAGAGCAGG	
258		2	+	0.19	ACAGCAAGAG	^	CAGGTTCCCC	
261		2	+	0.18	GCAAGAGCAG	^	GTTCCCCACA	
364		0	+	0.27	CCCATCCCAG	^	GATTCGATGC	
405		0	+	0.71	TTTTTCCAG	^	CGTCTGCACA	
416		2	+	0.00	GTCTGCACAG	^	TGCCAGGCAC	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
66		444		1	-	0.00	TTTATAAAAG	^	GTGCTCAGAG	

Desaparece el sitio *acceptor* (en rojo) en la secuencia mutante, mientras que el otro sitio aceptor (en verde) se modifica por la presencia de la mutación, pero su confianza no se ve alterado por lo que no se tendrá en cuenta. El sitio *acceptor* que se pierde no corresponde al sitio *acceptor* de *splicing* normal para este exón, por lo que perderlo no debería suponer ningún efecto para el *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.2.148.573458.0 :

Start	End	Score	Exon	Intron
29	43	0.51	cctgaag	gttggctg
323	337	0.81	gtgtcag	gtgggtgt
426	440	0.41	cagggag	gtgcttga

Donor site predictions for 10.42.3.123.573470.0 :

Start	End	Score	Exon	Intron
29	43	0.51	cctgaag	gttggctg
323	337	0.81	gtgtcag	gtgggtgt
426	440	0.41	cagggag	gtgcttga

Acceptor site predictions for 10.42.2.148.573458.0 :

Start	End	Score	Intron	Exon
97	137	0.98	ctggcctctccccctccctc	aggccaatgctaccgtcactct
135	175	0.92	tctctgacttttggtctttc	agataagttcacccaaaacccc
344	384	0.62	gatctgtcaccccatccc	aggattcgatgccaggggaagc
385	425	0.97	ccacgtctactttttttcc	agcgtctgcacagtgccaggca
462	502	0.68	ctcaggtctccttgattt	agggaacctcctcacaagttcc

Acceptor site predictions for 10.42.3.123.573470.0 :

Start	End	Score	Intron	Exon
97	137	0.98	ctggcctctccccctccctc	aggccaatgctaccgtcactct
135	175	0.92	tctctgacttttggtctttc	agataagttcacccaaaacccc
344	384	0.62	gatctgtcaccccatccc	aggattcgatgccaggggaagc
385	425	0.97	ccacgtctactttttttcc	agcgtctgcacagtgccaggca
462	502	0.68	ctcaggtctccttgattt	agggaacctcctcacaagttcc


Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
tcccc(c/a)cacag	cccccc	ccccac	35700	71%

CRYP-SKIP

Parece que hay un sitio críptico de *splicing* dentro del propio exón, pero el cambio de interés (la primera *a* en minúsculas detrás de las mayúsculas, que indican el exón) no lo toma en consideración, por lo que no debe considerar que tenga algún efecto en el *splicing*.

Human Splicing Finder

 No significant impact on splicing signals.

No significant impact on splicing signals.

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr			
wt	15	86	tcgtcagcc		-1.26850401419	0.506172839506	33	17	26	-1.7780551		
wt	15	71	gtgtcaggt		-2.47290537021	0.515151515152	18	17	26	-1.2972604		
wt	15	47	ctgtcaccc		0.675520099689	0.547619047619	1	10	17	0.93821187		
mut	15	86	tcgtcagcc		-1.26850401419	0.506172839506	33	17	26	-1.7780551		
mut	15	71	gtgtcaggt		-2.47290537021	0.515151515152	18	17	26	-1.2972604		
mut	15	47	ctgtcaccc		0.675520099689	0.547619047619	1	10	17	0.93821187		

Variant Effect Predictor tool

ENST00000263734.4:c.2401C>A	2:46382538-46382538	A	missense_variant	EPAS1	ENSG00000116016	Transcript	ENST00000263734.5	protein_coding	15/16	2906	2401	801	P/T	CCA/ACA	rs1265031766
ENST00000263734.4:c.2401C>A	2:46382538-46382538	A	downstream_gene_variant	EPAS1	ENSG00000116016	Transcript	ENST00000465318.5	retained_intron	-	-	-	-	-	-	rs1265031766
ENST00000263734.4:c.2401C>A	2:46382538-46382538	A	non_coding_transcript_exon_variant	EPAS1	ENSG00000116016	Transcript	ENST00000466465.5	retained_intron	4/5	1374	-	-	-	-	rs1265031766
ENST00000263734.4:c.2401C>A	2:46382538-46382538	A	non_coding_transcript_exon_variant	EPAS1	ENSG00000116016	Transcript	ENST00000468530.1	processed_transcript	3/4	317	-	-	-	-	rs1265031766
ENST00000263734.4:c.2401C>A	2:46382538-46382538	A	downstream_gene_variant	EPAS1	ENSG00000116016	Transcript	ENST00000483692.1	retained_intron	-	-	-	-	-	-	rs1265031766
ENST00000263734.4:c.2401C>A	2:46382538-46382538	A	regulatory_region_variant	-	-	RegulatoryFeature	ENSR00000604776	CTCF_binding_site	-	-	-	-	-	-	rs1265031766

ESEfinder

Se encuentran 2 predicciones que tienen puntuación positiva para las matrices 5' (247, 259) y una predicción que tiene puntuación positiva para las matrices 3' (260).

247 (-263)	GAACAGCAAGAGCAGGTTCCCCCAGTG	2.15740	247 (-263)	GAACAGCAAGAGCAGGTTCCCCCAGTG	-2.26920	247 (-263)	GAACAGCAAGAGCAGGTTCCCCCAGTG	2.35230	247 (-263)	GAACAGCAAGAGCAGGTTCCCCCAGTG	-3.05080
259 (-251)	CAGGTTCCCCCAGTGCTACGCCACCCA	0.14410	259 (-251)	CAGGTTCCCCCAGTGCTACGCCACCCA	-23.53210	259 (-251)	CAGGTTCCCCCAGTGCTACGCCACCCA	-0.01110	259 (-251)	CAGGTTCCCCCAGTGCTACGCCACCCA	-25.03860
260 (-250)	AGGTTCCCCCAGTGCTACGCCACCCAG	-30.00190	260 (-250)	AGGTTCCCCCAGTGCTACGCCACCCAG	7.61880	260 (-250)	AGGTTCCCCCAGTGCTACGCCACCCAG	-28.51450	260 (-250)	AGGTTCCCCCAGTGCTACGCCACCCAG	7.89490

Si comparamos estas puntuaciones con las equivalentes en la secuencia mutante comprobamos que para las matrices 5' han descendido las puntuaciones ligeramente, mientras que para 260 las puntuaciones han bajado más de 2 puntos.

247 (-263)	GAACAGCAAGAGCAGGTTCCCCACAGTG	1.75260	247 (-263)	GAACAGCAAGAGCAGGTTCCCCACAGTG	-2.31680	247 (-263)	GAACAGCAAGAGCAGGTTCCCCACAGTG	1.88490	247 (-263)	GAACAGCAAGAGCAGGTTCCCCACAGTG	-3.12500
259 (-251)	CAGGTTCCCCACAGTGCTACGCCACCCA	0.07570	259 (-251)	CAGGTTCCCCACAGTGCTACGCCACCCA	-25.64020	259 (-251)	CAGGTTCCCCACAGTGCTACGCCACCCA	-0.04030	259 (-251)	CAGGTTCCCCACAGTGCTACGCCACCCA	-27.29920
260 (-250)	AGGTTCCCCACAGTGCTACGCCACCCAG	-30.09590	260 (-250)	AGGTTCCCCACAGTGCTACGCCACCCAG	5.27340	260 (-250)	AGGTTCCCCACAGTGCTACGCCACCCAG	-28.54260	260 (-250)	AGGTTCCCCACAGTGCTACGCCACCCAG	5.49020

Por lo tanto, lo más probable es que se estuviera debilitando un sitio *acceptor*, que podría tener efecto en el *splicing*.

En cuanto a los sitios ESE, se ven algunos cambios entre ambas secuencias:

263 (-247)	TTCCCCC	-3.26185	263 (-247)	TTCCCCC	-1.61290	263 (-247)	TTCCCCC	0.84973	263 (-247)	TTCCCCC	0.73156
264 (-246)	TCCCCC	-3.86478	264 (-246)	TCCCCC	-1.86168	264 (-246)	TCCCCCA	1.27160	264 (-246)	TCCCCC	1.04362
265 (-245)	CCCCCA	1.29235	265 (-245)	CCCCCA	2.66776	265 (-245)	CCCCCAC	-0.89621	265 (-245)	CCCCCA	-0.85218
266 (-244)	CCCCAC	-0.91429	266 (-244)	CCCCAC	1.11102	266 (-244)	CCCCACA	0.01593	266 (-244)	CCCCAC	0.42471
267 (-243)	CCCCACA	-1.23138	267 (-243)	CCCCACA	0.75345	267 (-243)	CCCCACAG	0.89154	267 (-243)	CCCCACA	0.88480
268 (-242)	CCCACAG	1.55911	268 (-242)	CCCACAG	2.50585	268 (-242)	CCCACAGT	-6.16804	268 (-242)	CCCACAG	-0.30167
269 (-241)	CCACAGT	-1.31690	269 (-241)	CCACAGT	0.40803	269 (-241)	CCACAGTG	-0.01134	269 (-241)	CCACAGT	3.58597
263 (-247)	TTCCCCA	-1.05521	263 (-247)	TTCCCCA	-0.18851	263 (-247)	TTCCCCAC	0.47106	263 (-247)	TTCCCCA	-0.80754
264 (-246)	TCCCCAC	-3.86478	264 (-246)	TCCCCAC	-1.99403	264 (-246)	TCCCCACA	-0.00520	264 (-246)	TCCCCAC	0.78141
265 (-245)	CCCCACA	-1.23138	265 (-245)	CCCCACA	0.75345	265 (-245)	CCCCACAC	-1.36863	265 (-245)	CCCCACA	0.88480
266 (-244)	CCCACAC	0.07926	266 (-244)	CCCACAC	1.51683	266 (-244)	CCCACACA	-2.68221	266 (-244)	CCCACAC	-1.14431
267 (-243)	CCACACA	-3.54139	267 (-243)	CCACACA	-1.01157	267 (-243)	CCACACAG	-0.11914	267 (-243)	CCACACA	3.28035
268 (-242)	CACACAG	3.28457	268 (-242)	CACACAG	3.18589	268 (-242)	CACACAGT	-4.49282	268 (-242)	CACACAG	-2.56839
269 (-241)	ACACAGT	-3.82208	269 (-241)	ACACAGT	-2.72771	269 (-241)	ACACAGTG	0.26510	269 (-241)	ACACAGT	2.89305

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	1	3	2	15	158.4084	3	-8.3122	13	9	58	691.6558	62	86.3690	24	142	0.17
mut	1	3	2	15	158.4084	4	-9.3122	13	9	56	670.2345	62	86.7054	25	140	0.18

Allele mut has a higher chance of exon skipping than allele wt.

HOT-SKIP

```
tctgacttagatgatgccatcagagggtcctgaaggttggtgtagttctggtgactctgagcacctttataaaggaaa
gggatgctagggtcttcctggcctctccctccctcaggccaatgctaccgtcactctctgactttggtctttcagATAAG
TTACCCAAAACCCCATGAGGGGCTGGGCCATCCCCTGAGACATCTGCCGCTGCCACAGCCTCCATCTGCCATCAGTCC
CGGGGAGAACAGCAAGAGCAGGTTCCCACACAGTGCTACGCCACCCAGTACCAGGACTACAGCCTGTCTCAGCCCACA
AGGTGTCAggtgggtgtgcccaggatctgtcaccctccatcccaggattcgatgccagggaagcccacgtctactttttt
tccagcgtctgcacagtgccaggcacaggagggtgcttgacttgaagtcacctatacagggtcaggtctccttgattt
agggaacctcctcacaaagttccttgtagac
```