

Ejemplo comparación de resultados predictores in silico

Cambio de estudio FRMPD4 c.813+3C>G (chr23:12690329 C/G, COSV66215515 o NM_001368397.1: c.813+3C>G)

Exón 8 e intrones adyacentes:

```
ctttctatatcaagagccttacatgatattaagcttctcctccagctatttaccttgttaa
ctctctgttggtgaactccataaagattttgaatgatcgaggaggccaggaaaagggcaag
agaaggaaggcttcatttctgcatcttttttggatatttggagcattaaatttcagaatca
accaaacgtataaaaatttcaagattggcagcctacatgtgtaacaacttttcagtcatggg
agactccaaagatggcagccttcgatttttggtctcatcatgttcatgtttcctctgtaag
GATGTCATCTTAACCCTTCAAGAGAAGCTCTCCATCAAAGATTGAACACTTCTCTCTC
ATGCTGGAGCAAGAGGACAGAAGGGGCTGGAACGAAGCTGCTCTTCCTTCATGAAACAGGAG
ACTCTAACTCAG
gtctgtgaaatctcaccctccaagtgatggaggctgccaggttagagacagaggttgcccagtc
caagatttctcccccactctaattcctgtaaatgtattaattgtatcccctgggtatcctct
gtaaaggaagtgctggccatttcttaatgcatttttacacctcaaggtcatgtttccctttct
gcccatcttttgaagactgtcagtctcacaactagatttgtcattcaaagttctcaccagc
tgcacatttgagaatacagctctgtaaaattatattttatgggattacagtgaatctgttca
```

El cambio se encuentra en la tercera posición justo después del exón 8 (la **c** 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'
433		0	+	0.42	TCTAACTCAG	^GTCTGTGAAA		
470		1	+	0.66	GAGGCTGCAG	^GTTAGAGCAG		

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
668	65	2	-	0.00	TGTGCAGCTG	^GTGAGACTTT		
448	285	0	-	0.34	TCACCTTGAGG	^GTGAGATTTC		
52	681	1	-	0.00	AGTTAACAAAG	^GTAATAGCT		

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
300		0	+	0.19	CCTCTGTAAG	^GATGTCATCT		
322		1	+	0.32	ACCCCTCAAG	^AGAAGCTCTC		
324		0	+	0.19	CCTTCAAGAG	^AAGCTCTCCA		
327		0	+	0.18	TCAAGAGAAG	^CTCTCCATCA		
340		1	+	0.07	TCCATCAAAG	^GCATTGAACA		
627		1	+	0.16	TCTTTTGCAG	^ACTGTCAGTC		

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
551	182	1	-	0.43	TCCTTTACAG	^AGGATACCCA		
464	269	0	-	0.25	TAACCTGCAG	^CCTCATCACT		
397	336	0	-	0.07	GCAAGAGCAG	^CTTCGTTCCA		
386	347	2	-	0.18	TTCGTTCCAG	^CCCCTTCTGT		
364	369	0	-	0.85	TCTGCTCCAG	^CATGAGAGAG		
358	375	0	-	0.19	CCAGCATGAG	^AGAGAAGTGT		
356	377	2	-	0.19	AGCATGAGAG	^AGAAGTGTTT		
354	379	1	-	0.18	CATGAGAGAG	^AAGTGTTCAA		
351	382	1	-	0.17	GAGAGAGAAG	^TGTTCAATGC		

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
433		0	+	0.85	TCTAACTCAG	^GTGTGTGAAA		
470		1	+	0.63	GAGGCTGCAG	^GTTAGAGCAG		

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
668	65	2	-	0.00	TGTGCAGCTG	^GTGAGACTTT		
448	285	0	-	0.35	TCACCTTGAGG	^GTGAGATTTC		
52	681	1	-	0.00	AGTTAACAAAG	^GTAATAGCT		

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
300		0	+	0.19	CCTCTGTAAG	^GATGTCATCT		
322		1	+	0.33	ACCCCTCAAG	^AGAAGCTCTC		
324		0	+	0.19	CCTTCAAGAG	^AAGCTCTCCA		
327		0	+	0.18	TCAAGAGAAG	^CTCTCCATCA		
340		1	+	0.07	TCCATCAAAG	^GCATTGAACA		
627		1	+	0.16	TCTTTTGCAG	^ACTGTCAGTC		

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
551	182	1	-	0.43	TCCTTTACAG	^AGGATACCCA		
464	269	0	-	0.25	TAACCTGCAG	^CCTCATCACT		
386	347	2	-	0.18	TTCGTTCCAG	^CCCCTTCTGT		
364	369	0	-	0.85	TCTGCTCCAG	^CATGAGAGAG		
358	375	0	-	0.19	CCAGCATGAG	^AGAGAAGTGT		
356	377	2	-	0.19	AGCATGAGAG	^AGAAGTGTTT		
354	379	1	-	0.18	CATGAGAGAG	^AAGTGTTCAA		
351	382	1	-	0.17	GAGAGAGAAG	^TGTTCAATGC		

Los sitios *donor* en las dos secuencias son los mismos, a diferencia de la posición dónde se encuentra la mutación que pasa de ser una C a una G, haciendo que el sitio *donor* sea más estable (tiene más confianza que en la secuencia *wild type*) por lo que, a priori, no debería haber cambios en el *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.3.123.425571.0 :

Start	End	Score	Exon	Intron
463	477	0.87	gctgcag	gttagagc

Donor site predictions for 10.42.0.139.425583.0 :

Start	End	Score	Exon	Intron
426	440	0.91	aactcag	gtgtga
463	477	0.87	gctgcag	gttagagc

Acceptor site predictions for 10.42.3.123.425571.0 :

Start	End	Score	Intron	Exon
24	64	0.70	gatattaagcttctcctcc	agctatttaccttgtaactct
607	647	0.96	ctttctgcccacatcttttgc	agactgtcagtcacaaactag

Acceptor site predictions for 10.42.0.139.425583.0 :

Start	End	Score	Intron	Exon
24	64	0.70	gatattaagcttctcctcc	agctatttaccttgtaactct
607	647	0.96	ctttctgcccacatcttttgc	agactgtcagtcacaaactag

Aparece un nuevo sitio dolor en la secuencia mutante que tiene mayor score que el *donor* que predice en la secuencia *wild type*. El nuevo sitio *donor* que está prediciendo realmente es el sitio dolor del exón por lo que en teoría no debería haber cambios en el *splicing*.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
caggt(c/g)tgtga	caggtc	caggtg	29839	73%

Human Splicing Finder

<div> <div></div> <div>New Donor splice site</div> <div>Activation of a cryptic Donor site. Potential alteration of splicing</div> </div>			
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chrX:12690326	- REF : GGTCTGTGA - ALT : GGTGTGTGA	42.42 > 69.56 => 63.98%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr			
wt	14	78	ctctaactc		3.36791497834	0.479452054795	16	10	19	1.0395526		
wt	14	74	aactcaggt		-0.763764775782	0.463768115942	12	10	19	-0.33006743		
wt	14	65	ctgtgaaat		-0.724834184743	0.466666666667	3	10	19	0.25579677		
wt	14	58	atctcaccc		2.1212267767	0.452830188679	53	0	0	-1.9762217		
wt	14	52	ccctcaagt		0.355921405647	0.404255319149	47	0	0	-2.3033205		
wt	14	47	aagtgatga		-0.195519686717	0.428571428571	42	0	0	-2.1948908		
wt	14	44	tgatgaggc		-1.49700192963	0.435897435897	39	0	0	-2.5122215		
wt	14	32	aggttagag		-4.03870734181	0.444444444444	27	0	0	-2.745079		
mut	14	78	ctctaactc		3.36791497834	0.465753424658	16	10	19	1.0351284		
mut	14	74	aactcaggt		-0.763764775782	0.449275362319	12	10	19	-0.33474817		
mut	14	65	gtgtgaaat		-0.943029400712	0.466666666667	3	10	19	0.17036304		
mut	14	58	atctcaccc		2.1212267767	0.452830188679	53	0	0	-1.9762217		
mut	14	52	ccctcaagt		0.355921405647	0.404255319149	47	0	0	-2.3033205		
mut	14	47	aagtgatga		-0.195519686717	0.428571428571	42	0	0	-2.1948908		
mut	14	44	tgatgaggc		-1.49700192963	0.435897435897	39	0	0	-2.5122215		
mut	14	32	aggttagag		-4.03870734181	0.444444444444	27	0	0	-2.745079		

El cambio en las previsiones para ambas secuencias es que se sustituye la posición donde se encuentra la mutación de una C a una G, lo que provoca un descenso en la puntuación del BP, lo cual podría estar afectando al *splicing*.

Variant Effect Predictor tool

ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000380682.5	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000616992.1	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000640291.2	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000656302.1	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000657176.1	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant, non_coding_transcript_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000657982.1	processed_transcript	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000672010.1	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant, non_coding_transcript_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000672340.1	processed_transcript	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant, non_coding_transcript_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000672861.1	processed_transcript	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000672869.1	protein_coding	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant, non_coding_transcript_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000673271.1	processed_transcript	-	-	-	-	-	-	COSV66215515
ENST00000380682.5:c.813+3C>G	X-12690329-12690329	G	splice_region_variant, intron_variant	FRMPD4	ENSG00000169933	Transcript	ENST00000675598.1	protein_coding	-	-	-	-	-	-	COSV66215515

ESEfinder

Solo se obtiene un resultado con puntuaciones positivas para todas las matrices en la secuencia *wild type*. Cuando comparamos estas puntuaciones con las equivalentes en la secuencia mutante se observa que para las matrices 5' has subido considerablemente, mientras que para las 3' has descendido:

418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	6.02870	418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	2.07710	418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	5.33170	418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	1.86080
418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	9.64180	418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	1.97640	418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	8.96080	418 (-315)	GAGACTCTAACTCAGgtctgtgaaatctca	1.76190

Además se observan que una de las predicciones que tenía puntuación negativa para la secuencia *wild type* en la secuencia mutante pasan a ser positivas:

420 (-313)	GACTCTAACTCAGgtctgtgaaatctcacc	-9.01690	420 (-313)	GACTCTAACTCAGgtctgtgaaatctcacc	-32.05470	420 (-313)	GACTCTAACTCAGgtctgtgaaatctcacc	-9.43720	420 (-313)	GACTCTAACTCAGgtctgtgaaatctcacc	-31.44470
420 (-313)	GACTCTAACTCAGgtgtgtgaaatctcacc	1.74850	420 (-313)	GACTCTAACTCAGgtgtgtgaaatctcacc	-30.27430	420 (-313)	GACTCTAACTCAGgtgtgtgaaatctcacc	1.33090	420 (-313)	GACTCTAACTCAGgtgtgtgaaatctcacc	-29.63820

Por lo tanto se está modificando el sitio *donor* del *splicing* lo que afectara el proceso normal de este.