

Ejemplo comparación de resultados predictores in silico

Cambio de estudio PTPRT c.1865+1G>T (chr 20:42350627 G/T, COSV100625964 o NM_007050.6: c.1865+1G>T)

Exón 11 e intrones adyacentes:

```
t a g t a c t t a t a t t c c c t g c a c a t a t c t g t t t a t t a g g t g t g g a c g t t t t g g t t c a t g c a
g g t t a c t c g a c a g a t c a t t g c t a c c e t t c c c c a c a c t a c t g c a c c t t g a a g a a a c c t a c t
a t g g g c a g g t t g g g c a t c g g g g c c t c c t t t a c t g a a g g c o t g g a a g g c t t c t t a t c c a a c
c c c t t t c c t a a a c g t c c t c c t c t a t c c a t g c t g t g t c t t t a a g g a t g g c a g c a a g g g g c
g g t g g g a g c c a g t t g g g a a t c t g t g c t g a a c t c a c c t g c a c t c t c t g t t t c c t g t c c a g
C T C C A T C C A T G C C T G A G T A C G A C A C A G A C A C C C A T T G A A T G A G A C A G A C A C G A C C A T C A
C A G T G A T G C T G A A A C C C G C T C A G T C C C G G G G A G C T C T G T C A G
g t g a g g a t g g t t c t t c a c t t g g a g t c t g c a g t t t t t c t t c t c t g t g c c a c a t g g g c o t c c
a g t t g a a t c a t g t g a c t g g c a a t g c c c a t g g g c a a g a c t g g c c a g g a a c a a a g c t t c c t c
g g a g g a a g g a g g t c a g c a g c c c c a a g c c c a g t t c a a a g g g a g c c c t g a g t t c a a g g c a a
a c c t g g g t g a c a t a g g g a g a c c c c g t c t a t a c a a a g a a a a a a a t t t t t t a a a t t a g c t g g
a t g g g t a g c a c a c g c c t g t a g t a c c g g c t a c t c g a g a g g a t g a t a t g g a a a g a t t g c t t
```

El cambio se encuentra en la primera posición justo después del exón 11 (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'
404		2	+	0.93	CTCCTGTCAG	GTGAGGATGG	H	
535		0	+	0.45	GAGGAAGGAG	GTGAGGATGG		

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
276	428	1	-	0.99	GAGAGTGCAG	GTGAGTTCAG	H	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
36		2	+	0.00	TGTTTATTAG	GTGTGGAACG		
300		1	+	0.76	TCCTGTCCAG	CTCCATCCAT		
317		0	+	0.34	CATGCCTGAG	TACGACACAG		
327		1	+	0.25	TACGACACAG	ACACCCCAT		
465		0	+	0.18	GGGCCTCCAG	TTGAATCATG		

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
397	307	1	-	0.33	CACCTGACAG	GAGCTCCCCG		
313	391	1	-	0.30	TCGTAATCAG	GCATGGATGG		

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
535		0	+	0.44	GAGGAAGGAG	GTGAGGATGG		

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
276	428	1	-	0.99	GAGAGTGCAG	GTGAGTTCAG	H	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
36		2	+	0.00	TGTTTATTAG	GTGTGGAACG		
300		1	+	0.76	TCCTGTCCAG	CTCCATCCAT		
317		0	+	0.34	CATGCCTGAG	TACGACACAG		
327		1	+	0.20	TACGACACAG	ACACCCCAT		
465		0	+	0.18	GGGCCTCCAG	TTGAATCATG		

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
397	307	1	-	0.15	CAACTGACAG	GAGCTCCCCG		
313	391	1	-	0.33	TCGTAATCAG	GCATGGATGG		

Se está perdiendo un sitio *donor* (en azul) que se detectaba en la secuencia WT y no se predice en la secuencia mutante. Este coincide con el sitio *donor* del exón, por lo que se van a producir cambios en el *splicing*, bien saltándose el exón completo o bien (en el caso de que se empleara el *acceptor* normal y el otro sitio *donor* predicho) se incluirían 131 pares de bases en la secuencia del mensajero.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
55	69	0.54	catgcag	gttactcgc
397	411	0.99	ctgtcag	gtgaggat
528	542	0.53	gaaggag	gtcagcag

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
16	56	0.74	ctgcacatatctgtttatt	aggtgtggaacgttttggttca
204	244	0.71	tatccatgctgtgtcttta	aggatggcagcaagggcggtg
280	320	0.97	cactctctgtttcctgtcc	agctccatccatgcctgagtac

Donor site predictions for mut :

Start	End	Score	Exon	Intron
55	69	0.54	catgcag	gttactcgc
528	542	0.53	gaaggag	gtcagcag

Acceptor site predictions for mut :


Start	End	Score	Intron	Exon
16	56	0.74	ctgcacatatctgtttatt	aggtgtggaacgttttggttca
204	244	0.71	tatccatgctgtgtcttta	aggatggcagcaagggcggtg
280	320	0.97	cactctctgtttcctgtcc	agctccatccatgcctgagtac

Se está perdiendo un sitio *donor* (en azul) que se detectaba en la secuencia WT y no se predice en la secuencia mutante. Este coincide con el sitio *donor* del exón, por lo que se van a producir cambios en el *splicing*, bien saltándose el exón completo o bien (en el caso de que se empleara el *acceptor* normal y el siguiente sitio *donor* predicho) se incluirían 131 pares de bases en la secuencia del mensajero.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gtcag(g/t)tgagg	caggtg	cagttg	29806	73%

Human Splicing Finder

 Broken WT Donor Site	Alteration of the WT Donor site, most probably affecting splicing		
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr20:42350630	- REF : CAGGTGAGG - ALT : CAGTTGAGG	94.56 > 67.42 => -28.7%
MaxEnt Donor site	chr20:42350630	- REF : CAGGTGAGG - ALT : CAGTTGAGG	10.07 > 1.56 => -84.51%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	16	65	ccatcacag		-0.902916729791	0.5	60	0	-3.5881706
wt	16	59	cagtgatgc		1.56505381791	0.518518518519	54	0	-2.2360726
wt	16	53	tgctgaaac		1.5325273915	0.520833333333	48	0	-1.8682707
wt	16	43	cgctcagtc		1.55421593035	0.5	38	0	-1.233524
wt	16	23	ctgtcaggt		-2.25471015424	0.444444444444	18	0	-1.476875
wt	16	18	aggtgagga		-2.33930154855	0.538461538462	13	0	-1.1631401
mut	16	65	ccatcacag		-0.902916729791	0.516666666667	60	0	-3.5827878
mut	16	59	cagtgatgc		1.56505381791	0.537037037037	54	0	-2.2300917
mut	16	53	tgctgaaac		1.5325273915	0.541666666667	48	0	-1.8615422
mut	16	43	cgctcagtc		1.55421593035	0.526315789474	38	0	-1.2250248
mut	16	23	ctgtcagtt		-1.86300009501	0.5	18	0	-1.3055592
mut	16	18	agttgagga		-2.4030353825	0.538461538462	13	0	-1.1880949

Se producen cambios en los dos BP debido a la presencia de la mutación que aumentan el score pero este sigue siendo negativo, por lo tanto no se van a tener en cuenta.

Variant Effect Predictor tool

ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000356100.6	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000373184.5	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000373187.5	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000373190.5	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000373193.7	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000373198.8	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000373201.5	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000612229.4	protein_coding	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant, NMD_transcript_variant	PTPRT	ENSG00000196090	Transcript	ENST00000617474.1	nonsense_mediated_decay	-	-	-	-	-	-	-	COSV100625964
ENST00000373193.7:c.1865+1G>T	20:42350627-42350627	A	splice_donor_variant	PTPRT	ENSG00000196090	Transcript	ENST00000618610.1	protein_coding	-	-	-	-	-	-	-	COSV100625964

ESEfinder

Se encuentran dos predicciones con puntuaciones positivas: una solo para las matrices 5' (384) y una con puntuaciones positivas para todas las matrices (389).

384 (-320)	TCCCGGGGAGCTCCTGTCAGgtgaggatgg	1.91450	384 (-320)	TCCCGGGGAGCTCCTGTCAGgtgaggatgg	-25.95170	384 (-320)	TCCCGGGGAGCTCCTGTCAGgtgaggatgg	1.86510	384 (-320)	TCCCGGGGAGCTCCTGTCAGgtgaggatgg	-24.99210
389 (-315)	GGGAGCTCCTGTCAGgtgaggatggttctt	11.82000	389 (-315)	GGGAGCTCCTGTCAGgtgaggatggttctt	2.94240	389 (-315)	GGGAGCTCCTGTCAGgtgaggatggttctt	11.67660	389 (-315)	GGGAGCTCCTGTCAGgtgaggatggttctt	2.84820

Si comparamos estas puntuaciones con las de la secuencia mutante, se observa que las puntuaciones para 384 han aumentado, mientras que para 389 disminuido en todas las matrices.

384 (-320)	TCCCGGGGAGCTCCTGTCAGttgaggatgg	2.86270	384 (-320)	TCCCGGGGAGCTCCTGTCAGttgaggatgg	-26.48270	384 (-320)	TCCCGGGGAGCTCCTGTCAGttgaggatgg	2.84320	384 (-320)	TCCCGGGGAGCTCCTGTCAGttgaggatgg	-25.50050
389 (-315)	GGGAGCTCCTGTCAGttgaggatggttctt	-5.18450	389 (-315)	GGGAGCTCCTGTCAGttgaggatggttctt	0.24980	389 (-315)	GGGAGCTCCTGTCAGttgaggatggttctt	-5.67120	389 (-315)	GGGAGCTCCTGTCAGttgaggatggttctt	0.23700

Por lo tanto, lo más probable es que se esté perdiendo un *donor* en la secuencia mutante, ya que la puntuación de 389 en 5' disminuye en casi 17 puntos.