Ejemplo comparación de resultados predictores in sillico

Cambio de estudio TP53 c.375+1G>A (chr17:7675993 G/A, rs1567555445o NM_000546.6: c.375+1G>A)

Exón 4 e intrones adyacentes:

```
gtaaggacaaggggttgggctggggacctggagggctggggacctggagggctgggggggct
ggggggctgaggacctggtcctctgactgctctttcacccatctacag
TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA
TGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCCC
GTGGCCCCTGCACCAGCAGCTCCTACACCGGCGGCCCCTGCACCAGCCCCCTCCTGGCCC
CTGTCATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGC
TTCTTGCATTCTGGGACAGCCAAGTCTGTGACTTGCACG
gtcagttgccctgaggggctggcttccatgagacttcaatgcctggccgtatccccctgc
atttcttttgtttggaactttgggattcctcttcacccttggcttcctgtcagtgtttt
tttatagtttacccacttaatgtgtgatctctgactcctgtcccaaagttgaatattccc
ccccttgaatttgggctttttatccatcccatcacaccctcagcatctctcctggggatgca
gaacttttctttttcttcatccacgtgtattccttggcttttgaaaataagctcctgacc
aggcttggtggctcacacctgcaatcccagcactctcaaagaggccaaggcaggcagatc
acctgagcccaggagttcaagaccagcctgggtaacatgatgaaacctcgtctctacaaa
```

El cambio se encuentra en primera posición justo después del exón 4 (la **g** 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'
                   389       0     +        0.70       GACTTGCACG^GTCAGTTGCC

Donor splice sites, complement strand
-------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'      exon intron    3'
       322         487       2     -        0.71       AGCTGCCCTG^GTAGGTTTTC

Acceptor splice sites, direct strand
------------------------------------
               pos 5'->3'  phase strand  confidence  5'    intron exon      3'
                   109       0     +        0.96       CCATCTACAG^TCCCCCTTGC
                   128       1     +        0.25       CCGTCCCAAG^CAATGGATGA
                   313       0     +        0.93       CCCTTCCCAG^AAAACCTACC
                   325       0     +        0.07       AACCTACCAG^GGCAGCTACG
                   609       2     +        0.15       ACACCCTCAG^CATCTCTCCT

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon      3'
       360         449       0     -        0.64       GCTGTCCCAG^AATGCAAGAA
        43         766       0     -        0.00       AGCCCTCCAG^GTCCCCAGCC
```

```
Donor splice sites, direct strand
---------------------------------
   No donor site predictions above threshold.

Donor splice sites, complement strand
-------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'      exon intron    3'
       322         487       2     -        0.71       AGCTGCCCTG^GTAGGTTTTC

Acceptor splice sites, direct strand
------------------------------------
               pos 5'->3'  phase strand  confidence  5'    intron exon      3'
                   109       0     +        0.96       CCATCTACAG^TCCCCCTTGC
                   128       1     +        0.25       CCGTCCCAAG^CAATGGATGA
                   313       0     +        0.93       CCCTTCCCAG^AAAACCTACC
                   325       0     +        0.07       AACCTACCAG^GGCAGCTACG
                   609       2     +        0.15       ACACCCTCAG^CATCTCTCCT

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon      3'
       360         449       0     -        0.65       GCTGTCCCAG^AATGCAAGAA
        43         766       0     -        0.00       AGCCCTCCAG^GTCCCCAGCC
```

Se está perdiendo un sitio *donor* en la secuencia mutante. Este coincide con el sitio *donor* del exón 4, por lo que estará afectando al *splicing*, probablemente perdiéndose el exón completo ya que no se detecta otro sitio *donor* en la hebra directa.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.0.139.315469.0 :

Start	End	Score	Exon	Intron
382	396	0.59	ttgcacg	gtcagttg

Donor site predictions for 10.42.3.123.315859.0 :

Start	End	Score	Exon	Intron

Acceptor site predictions for 10.42.0.139.315469.0 :

Start	End	Score	Intron	Exon
89	129	0.64	gctcttttcacccatctac	agtccccttgccgtcccaagc
293	333	0.89	tcatcttctgtcccttccc	agaaaacctaccagggcagcta
495	535	0.99	cctgtcagtgtttttttat	agtttacccacttaatgtgtga

Acceptor site predictions for 10.42.3.123.315859.0 :

Start	End	Score	Intron	Exon
89	129	0.64	gctcttttcacccatctac	agtccccttgccgtcccaagc
293	333	0.89	tcatcttctgtcccttccc	agaaaacctaccagggcagcta
495	535	0.99	cctgtcagtgtttttttat	agtttacccacttaatgtgtga

Se está perdiendo un sitio *donor* en la secuencia mutante. Este coincide con el sitio *donor* del exón 4, por lo que estará afectando al *splicing*, probablemente perdiéndose el exón completo ya que no se detecta otro sitio *donor* en la hebra directa.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gcacg(g/a)tcagt	ggtcag	gatcag	31909	52%

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)	chr17:7675996	- REF : ACGGTCAGT - ALT : ACGATCAGT	87.47 > 60.33 => -31.03%
MaxEnt Donor site	chr17:7675996	- REF : ACGGTCAGT - ALT : ACGATCAGT	8.02 > -0.16 => -102%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	37	78	ctgtgactt	1.28093191374	0.602739726027	58	16	26	-2.3311001
wt	37	66	cggtcagtt	-1.00825174026	0.606557377049	46	16	26	-2.4666109
wt	37	56	ccctgaggg	0.492741774974	0.607843137255	36	16	26	-1.2455025
wt	37	39	ccatgagac	-0.948831341482	0.676470588235	19	16	26	-0.71171041
wt	37	32	acttcaatg	-2.14672886083	0.703703703704	12	16	26	-0.72886012
mut	37	78	ctgtgactt	1.28093191374	0.602739726027	58	16	26	-2.3311001
mut	37	66	cgatcagtt	-2.03508075528	0.606557377049	46	16	26	-2.868663
mut	37	56	ccctgaggg	0.492741774974	0.607843137255	36	16	26	-1.2455025
mut	37	39	ccatgagac	-0.948831341482	0.676470588235	19	16	26	-0.71171041
mut	37	32	acttcaatg	-2.14672886083	0.703703703704	12	16	26	-0.72886012

El cambio que se da entre ambas secuencias está en el BP que contiene la posición de interés. Sin embargo, como las puntuaciones son negativas, no se va a tener en cuenta.

Variant Effect Predictor tool

ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000269305.9	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000359597.8	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000413465.6	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000420246.6	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000445888.6	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000455263.6	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000503591.1	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	upstream_gene_variant	TP53	ENSG00000141510	Transcript	ENST00000504290.5	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	upstream_gene_variant	TP53	ENSG00000141510	Transcript	ENST00000504937.5	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant, non_coding_transcript_variant	TP53	ENSG00000141510	Transcript	ENST00000505014.5	retained_intron	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	splice_donor_variant	TP53	ENSG00000141510	Transcript	ENST00000508793.5	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628
ENST00000445888.6:c.375+1G>A	17:7675993-7675993	T	intron_variant	TP53	ENSG00000141510	Transcript	ENST00000509690.5	protein_coding	-	-	rs1567555445, CS951538, COSV52663569, COSV52668477, COSV52702628

ESEfinder

Para los resultados dónde se encuentra la posición de interés, observamos que se obtienen dos resultados con puntuación positivapara las matrices 5' (362 y 374) y un único resultado para las matrices 3' con puntuación positiva (379):

362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGgtc	2.17840	362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGgtc	-30.94350	362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGgtc	2.34700	362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGgtc	-29.97490
374 (-435)	TCTGTGACTTGCACGgtcagttgccctgag	5.64780	374 (-435)	TCTGTGACTTGCACGgtcagttgccctgag	-13.25930	374 (-435)	TCTGTGACTTGCACGgtcagttgccctgag	5.72280	374 (-435)	TCTGTGACTTGCACGgtcagttgccctgag	-11.59380
379 (-430)	GACTTGCACGgtcagttgccctgaggggct	-12.30430	379 (-430)	GACTTGCACGgtcagttgccctgaggggct	1.28560	379 (-430)	GACTTGCACGgtcagttgccctgaggggct	-12.54740	379 (-430)	GACTTGCACGgtcagttgccctgaggggct	0.87710

Si observamos las predicciones equivalentes para la secuencia mutante, se observa que la predicción positiva para las matrices 3’ (379) sigue siendo positiva igual que una de las predicciones para la matriz 5’ (362). Por el contrario, la predicción 374 tiene puntuaciones negativas donde antes eran positivas, por lo tanto se está perdiendo un sitio *donor* en la secuencia mutante, lo que llevara cambios en el *splicing:*

362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGatc	1.73910	362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGatc	-31.30600	362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGatc	1.81350	362 (-447)	GGGACAGCCAAGTCTGTGACTTGCACGatc	-30.30250
374 (-435)	TCTGTGACTTGCACGatcagttgccctgag	-5.33470	374 (-435)	TCTGTGACTTGCACGatcagttgccctgag	-14.58930	374 (-435)	TCTGTGACTTGCACGatcagttgccctgag	-5.40550	374 (-435)	TCTGTGACTTGCACGatcagttgccctgag	-12.84420
379 (-430)	GACTTGCACGatcagttgccctgaggggct	-12.07620	379 (-430)	GACTTGCACGatcagttgccctgaggggct	1.43070	379 (-430)	GACTTGCACGatcagttgccctgaggggct	-12.28100	379 (-430)	GACTTGCACGatcagttgccctgaggggct	0.96920