Ejemplo comparación de resultados predictores in sillico

Cambio de estudio CLIP4 c.601G>A (chr2:29145268 G/A, COSV60750480 o NM_024692.6: c.601G>A)
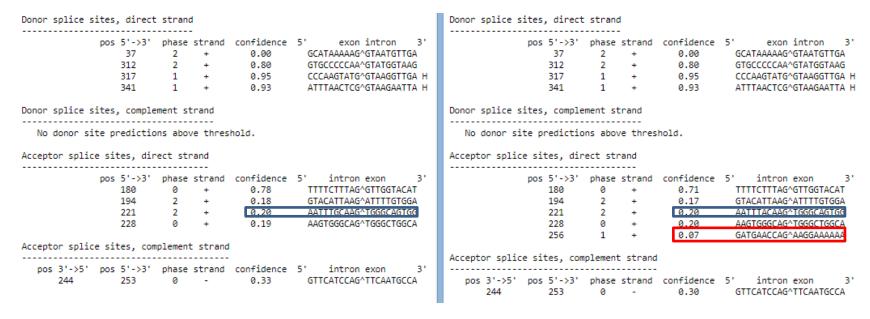
Exón 7 e intrones adyacentes:

El cambio se encuentra en la primera línea del exón 7 (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                          Donor splice sites, direct strand
---------------------------------                          ---------------------------------
        pos 5'->3' phase strand confidence 5'     exon intron    3'        pos 5'->3' phase strand confidence 5'     exon intron    3'
              37     2     +        0.00       GCATAAAAAG^GTAATGTTGA               37     2     +        0.00       GCATAAAAAG^GTAATGTTGA
             312     2     +        0.80       GTGCCCCCAA^GTATGGTAAG              312     2     +        0.80       GTGCCCCCAA^GTATGGTAAG
             317     1     +        0.95       CCCAAGTATG^GTAAGGTTGA H            317     1     +        0.95       CCCAAGTATG^GTAAGGTTGA H
             341     1     +        0.93       ATTTAACTCG^GTAAGAATTA H            341     1     +        0.93       ATTTAACTCG^GTAAGAATTA H

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

Acceptor splice sites, direct strand                       Acceptor splice sites, direct strand
------------------------------------                       ------------------------------------
        pos 5'->3' phase strand confidence 5'   intron exon    3'        pos 5'->3' phase strand confidence 5'   intron exon    3'
             180     0     +        0.78       TTTTCTTTAG^GTTGGTACAT              180     0     +        0.71       TTTTCTTTAG^GTTGGTACAT
             194     2     +        0.18       GTACATTAAG^ATTTTGTGGA              194     2     +        0.17       GTACATTAAG^ATTTTGTGGA
             221     2     +        0.20       AATTTGCAAG^TGGGCAGTGG              221     2     +        0.20       AATTTACAAG^TGGGCAGTGG
             228     0     +        0.19       AAGTGGGCAG^TGGGCTGGCA              228     0     +        0.20       AAGTGGGCAG^TGGGCTGGCA
                                                                                 256     1     +        0.07       GATGAACCAG^AAGGAAAAAA

Acceptor splice sites, complement strand                   Acceptor splice sites, complement strand
----------------------------------------                   ----------------------------------------
  pos 3'->5'  pos 5'->3'  phase strand confidence 5'   intron exon    3'    pos 3'->5'  pos 5'->3'  phase strand confidence 5'   intron exon    3'
     244         253        0     -        0.33       GTTCATCCAG^TTCAATGCCA        244         253        0     -        0.30       GTTCATCCAG^TTCAATGCCA
```

Uno de los sitios *acceptor* predichos (en azul) cambia por la presencia de la mutación, pero como está dentro del exón no se tendrá en cuenta. Aparece un nuevo sitio *acceptor* (en rojo) en la secuencia mutante. Tiene poca confianza pero, si se emplear, se perderían los primeros 76 nt del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon Intron
30	44	0.62	taaaaag**gt**aatgtt
174	188	0.75	tctttag**gt**tggtac
310	324	0.99	aagtatg**gt**aaggtt
334	348	1.00	taactcg**gt**aagaat

Donor site predictions for mut :

Start	End	Score	Exon Intron
30	44	0.62	taaaaag**gt**aatgtt
174	188	0.75	tctttag**gt**tggtac
310	324	0.99	aagtatg**gt**aaggtt
334	348	1.00	taactcg**gt**aagaat

Acceptor site predictions for wt :

Start	End	Score	Intron Exon
160	200	0.99	tctggtttatattttcttt**ag**gttggtacattaagattttg
362	402	0.90	tcaagttttactcttttac**ag**ttgttaaataaaacttttct

Acceptor site predictions for mut :

Start	End	Score	Intron Exon
160	200	0.99	tctggtttatattttcttt**ag**gttggtacattaagattttg
362	402	0.90	tcaagttttactcttttac**ag**ttgttaaataaaacttttct

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
aattt(g/a)caagt	aatttg	aattta	26790	54%

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      20    57       acattaaga  -4.42217522307  0.403846153846  4        6        13       -1.3313696
wt      20    56       cattaagat  -0.888829730904 0.411764705882  3        6        13       0.11795799
wt      20    37       aactgaatt  0.166028747425  0.40625 32      0        0        -1.4275535
mut     20    57       acattaaga  -4.42217522307  0.403846153846  4        6        13       -1.3313696
mut     20    56       cattaagat  -0.888829730904 0.411764705882  3        6        13       0.11795799
mut     20    37       aactgaatt  0.166028747425  0.40625 32      0        0        -1.4275535
mut     20    32       aatttacaa  -2.83397629178  0.37037037037   27       0        0        -2.2972937
```

Aparece un nuevo BP en la secuencia mutante pero, como tiene puntuación negativa, no se va a tener en cuenta.

Variant Effect Predictor tool

ENST00000401617.6:c.601G>A	2:29145268-29145268	A	missense_variant	CLIP4	ENSG00000115295 Transcript	ENST00000320081.10 protein_coding	8/16	1188	922	308	A/T	GCA/ACA	COSV60750480
ENST00000401617.6:c.601G>A	2:29145268-29145268	A	missense_variant	CLIP4	ENSG00000115295 Transcript	ENST00000401605.5 protein_coding	8/15	1180	922	308	A/T	GCA/ACA	COSV60750480
ENST00000401617.6:c.601G>A	2:29145268-29145268	A	missense_variant	CLIP4	ENSG00000115295 Transcript	ENST00000401617.6 protein_coding	7/16	1009	601	201	A/T	GCA/ACA	COSV60750480
ENST00000401617.6:c.601G>A	2:29145268-29145268	A	missense_variant	CLIP4	ENSG00000115295 Transcript	ENST00000404424.5 protein_coding	8/16	1254	922	308	A/T	GCA/ACA	COSV60750480
ENST00000401617.6:c.601G>A	2:29145268-29145268	A	missense_variant, NMD_transcript_variant	CLIP4	ENSG00000115295 Transcript	ENST00000415891.5 nonsense_mediated_decay	8/15	1144	922	308	A/T	GCA/ACA	COSV60750480
ENST00000401617.6:c.601G>A	2:29145268-29145268	A	3_prime_UTR_variant, NMD_transcript_variant	CLIP4	ENSG00000115295 Transcript	ENST00000456385.5 nonsense_mediated_decay	7/14	1115	-	-	-	-	COSV60750480

ESEfinder

Se encuentran dos resultados con puntuaciones positivas para 5'SS (206 y 213). Cuando se comparan las puntuaciones con las equivalentes en la secuencia mutante, estas han aumentado muy poco, por lo que no se tendrán en cuenta.

Fila superior (secuencia TTGCAAG)

Pos	Secuencia	Score 1	Score 2	Score 3	Score 4
206 (-291)	CAACTGAATTTGCAAGTGGGCAGTGGGCTG	6.01150	-8.07500	6.43490	-12.11260
207 (-290)	AACTGAATTTGCAAGTGGGCAGTGGGCTGG	-28.56060	-3.00500	-26.59910	-3.83690
208 (-289)	ACTGAATTTGCAAGTGGGCAGTGGGCTGGC	-17.97980	-28.19760	-15.99430	-27.73130
209 (-288)	CTGAATTTGCAAGTGGGCAGTGGGCTGGCA	-10.61790	-15.55190	-8.50330	-17.29640
210 (-287)	TGAATTTGCAAGTGGGCAGTGGGCTGGCAT	-5.76990	-15.22540	-6.00260	-13.64570
211 (-286)	GAATTTGCAAGTGGGCAGTGGGCTGGCATT	-22.23230	-25.34650	-18.72720	-24.37280
212 (-285)	AATTTGCAAGTGGGCAGTGGGCTGGCATTG	-31.30250	-32.88780	-29.28130	-31.83190
213 (-284)	ATTTGCAAGTGGGCAGTGGGCTGGCATTGA	0.61860	-36.37090	0.54880	-38.73780

Fila inferior (secuencia TTACAAG)

Pos	Secuencia	Score 1	Score 2	Score 3	Score 4
206 (-291)	CAACTGAATTTACAAGTGGGCAGTGGGCTG	6.19730	-8.14620	6.57470	-12.11120
207 (-290)	AACTGAATTTACAAGTGGGCAGTGGGCTGG	-28.33250	-2.85990	-26.33270	-3.74480
208 (-289)	ACTGAATTTACAAGTGGGCAGTGGGCTGGC	-18.07310	-28.11040	-16.06560	-27.57500
209 (-288)	CTGAATTTACAAGTGGGCAGTGGGCTGGCA	-10.86130	-15.55130	-8.74060	-17.39610
210 (-287)	TGAATTTACAAGTGGGCAGTGGGCTGGCAT	-5.72910	-15.51690	-5.92520	-13.99510
211 (-286)	GAATTTACAAGTGGGCAGTGGGCTGGCATT	-22.38880	-25.84030	-18.88650	-25.05620
212 (-285)	AATTTACAAGTGGGCAGTGGGCTGGCATTG	-31.46550	-33.51390	-29.48630	-32.63140
213 (-284)	ATTTACAAGTGGGCAGTGGGCTGGCATTGA	0.67520	-36.99090	0.61560	-39.47680

En cuanto a los ESE (desde 81 a 87), no se producen alteraciones que pueden estar afectando al *splicing:*

Pos	Seq	Value
77 (-420)	gagtgaa	-0.86717
78 (-419)	agtgaat	-7.03662
79 (-418)	gtgaatt	-1.35397
80 (-417)	tgaattt	-4.11563
81 (-416)	gaatttt	-4.74610
82 (-415)	aattttt	-5.67398
83 (-414)	attttta	-6.44326
84 (-413)	tttttaa	-7.34627
85 (-412)	ttttaaa	-7.34627
86 (-411)	tttaaaa	-4.89540
87 (-410)	ttaaaaa	-4.89540

Pos	Seq	Value
81 (-416)	gaatttt	-4.50566
82 (-415)	aattttt	-5.04705
83 (-414)	attttta	-5.25061
84 (-413)	tttttaa	-5.45626
85 (-412)	ttttaaa	-5.55879
86 (-411)	tttaaaa	-3.89823
87 (-410)	ttaaaaa	-3.59439
88 (-409)	taaaaac	-4.58753
89 (-408)	aaaaact	-3.28917
90 (-407)	aaaactc	-2.46756
91 (-406)	aaactcc	-4.78915

Pos	Seq	Value
77 (-420)	gagtgaat	-1.90536
78 (-419)	agtgaatt	-2.57687
79 (-418)	gtgaattt	-3.66333
80 (-417)	tgaatttt	-4.05523
81 (-416)	gaattttt	0.10086
82 (-415)	aattttta	0.60081
83 (-414)	atttttaa	-0.70913
84 (-413)	tttttaaa	-0.55448
85 (-412)	ttttaaaa	-0.26765
86 (-411)	tttaaaaa	-2.73495
87 (-410)	ttaaaaac	-4.98697

Pos	Seq	Value
81 (-416)	gaatttt	-5.48129
82 (-415)	aattttt	-5.07385
83 (-414)	attttta	-3.11919
84 (-413)	tttttaa	-0.61485
85 (-412)	ttttaaa	-0.36912
86 (-411)	tttaaaa	0.44648
87 (-410)	ttaaaaa	1.49376
88 (-409)	taaaaac	1.07820
89 (-408)	aaaaact	-1.24831
90 (-407)	aaaactc	-3.16312
91 (-406)	aaactcc	1.61409

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	3	2	1	34	557.1484	16	-28.1152	9	28	58	811.4881	52	62.9019	56	147	0.38
mut	3	2	1	31	521.5856	19	-30.0118	10	29	62	860.3316	53	63.1573	56	154	0.36

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

HOT-SKIP

gaagtgagagcaaggaacattaccaagcataaaaaggtaatgttgagagagcgaatttttaaaattcactttgagagagt
gaatttttaaaaactcccatggtgaagtaaagttgcaagagtaggaccccttggaattactaataattcccaaaattatt
ctggtttatattttctttagGTTGGTACATTAAGATTTTGTGGAACAACTGAATTTGCAAGTGGGCAGTGGGCTGGCATT
GAACTGGATGAACCAGAAGGAAAAAATAATGGAAGTGTTGGAAAAGTCCAGTACTTTAAATGTGCCCCCAAGTATGgtaa
ggttgatattatttaactcggtaagaattaattaaaaataatcaagtttactcttttacagttgttaaataaactttt
ctcctgatttcttatgtgataaatgaggggcatgtggatgtagggaggactttttaaagataggagatagttatatttat
ttacactgacgagagg

141	37	G	TGAATTTGCAAGTGG	AATTTGCAAGT	0	0	0	3	35.5628	0	-1.3482	0	0	2	15.3710	1	1.7226	3	3	1.00
142	37	A	TGAATTTACAAGTGG	AATTTACAAGT	0	0	0	0	0.0000	3	-3.2448	1	1	6	64.2144	2	1.9780	3	10	0.30
143	37	C	TGAATTTCCAAGTGG	AATTTCCAAGT	0	0	0	0	0.0000	0	-1.6028	0	0	0	0.0000	1	1.4680	0	1	0.00
144	37	T	TGAATTTTCAAGTGG	AATTTTCAAGT	0	0	0	3	31.1173	3	-3.2549	0	0	2	15.7162	0	1.2991	6	2	3.00