Ejemplo comparación de resultados predictores in sillico

Cambio de estudio KRAS c.34G>C (chr12:25245351 G/C, rs121913530 o NM_033360.4: c.34G>C)

Exón 2 e intrones adyacentes:

El cambio se encuentra en la primera línea del exón 2 (la **g** en color rojo subrayada de 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                              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'
            79        0      +        0.00       TCATATAAAG^GTGAGTTTGT            79        0      +        0.00       TCATATAAAG^GTGAGTTTGT
            97        1      +        0.37       GTATTAAAAG^GTACTGGTGG            97        1      +        0.37       GTATTAAAAG^GTACTGGTGG
           303        0      +        0.65       AACAATAGAG^GTAAATCTTG           303        0      +        0.69       AACAATAGAG^GTAAATCTTG

Donor splice sites, complement strand                          Donor splice sites, complement strand
-----------------------------------                            -----------------------------------
  pos 3'->5'  pos 5'->3'  phase strand  confidence 5'    exon intron   3'    pos 3'->5'  pos 5'->3'  phase strand  confidence 5'    exon intron   3'
    434          49        2      -        0.00       GAAACCCAAG^GTACATTTCA      434          49        2      -        0.00       GAAACCCAAG^GTACATTTCA

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        1      +        0.18       TTATTATAAG^GCCTGCTGAA          180        1      +        0.17       TTATTATAAG^GCCTGCTGAA
           338        1      +        0.19       ACTGGTGCAG^GACCATTCTT          216        1      +        0.07       CTTGTGGTAG^TTGGAGCTCG
                                                                               222        1      +        0.07       GTAGTTGGAG^CTCGTGGCGT
Acceptor splice sites, complement strand                                       234        1      +        0.07       CGTGGCGTAG^GCAAGAGTGC
-----------------------------------                                            338        1      +        0.18       ACTGGTGCAG^GACCATTCTT
    No acceptor site predictions above threshold.
                                                               Acceptor splice sites, complement strand
-----------------------------------                            -----------------------------------
                                                                   No acceptor site predictions above threshold.
```

Aparecen tres sitios *acceptor* nuevos (en azul) en la secuencia mutante. Tienen poca confianza, pero, si el *spliceosome* los reconociera, se produciría la pérdida de los primeros 36, 42 o 54 nt del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
72	86	0.99	tataaag**gt**gagttt	
90	104	0.79	ttaaaag**gt**actggt	
296	310	0.99	aatagag**gt**aaatct	

Donor site predictions for mut :

Start	End	Score	Exon	Intron
72	86	0.99	tataaag**gt**gagttt	
90	104	0.79	ttaaaag**gt**actggt	
296	310	0.99	aatagag**gt**aaatct	

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
160	200	0.69	tttcattatttttattata**ag**gcctgctgaaaatgactgaa	
336	376	0.67	caggaccattctttgatac**ag**ataaaggtttctctgaccat	

Acceptor site predictions for mut :

Start	End	Score	Intron	Exon
160	200	0.69	tttcattatttttattata**ag**gcctgctgaaaatgactgaa	
336	376	0.67	caggaccattctttgatac**ag**ataaaggtttctctgaccat	

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gagct(g/c)gtggc	agctgg	agctcg	31152	81%

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      12    65       tgctgaaaa  0.481644311634 0.383333333333 60       0        0        -3.0837295
wt      12    59       aaatgactg  1.10723640726  0.407407407407 54       0        0        -2.4512154
wt      12    55       gactgaata  0.297230706651 0.4            50       0        0        -2.51757
wt      12    49       atataaact  -1.22794910316 0.409090909091 44       0        0        -2.732024
mut     12    65       tgctgaaaa  0.481644311634 0.4            60       0        0        -3.0783466
mut     12    59       aaatgactg  1.10723640726  0.425925925926 54       0        0        -2.4452345
mut     12    55       gactgaata  0.297230706651 0.42           50       0        0        -2.5111106
mut     12    49       atataaact  -1.22794910316 0.431818181818 44       0        0        -2.7246838
```

Variant Effect Predictor tool

ENST00000256078.8:c.34G>C	12:25245351-25245351	G	▌missense_variant	KRAS	ENSG00000133703 Transcript	ENST00000256078.10 protein_coding	2/6	224	34	12	G/R	GGT/CGT	rs121913530, CM076251, COSV55497461, COSV55497469, COSV55497582, COSV56157736
ENST00000256078.8:c.34G>C	12:25245351-25245351	G	▌missense_variant	KRAS	ENSG00000133703 Transcript	ENST00000311936.8 protein_coding	2/5	224	34	12	G/R	GGT/CGT	rs121913530, CM076251, COSV55497461, COSV55497469, COSV55497582, COSV56157736
ENST00000256078.8:c.34G>C	12:25245351-25245351	G	▌missense_variant	KRAS	ENSG00000133703 Transcript	ENST00000556131.1 protein_coding	2/3	211	34	12	G/R	GGT/CGT	rs121913530, CM076251, COSV55497461, COSV55497469, COSV55497582, COSV56157736
ENST00000256078.8:c.34G>C	12:25245351-25245351	G	▌missense_variant	KRAS	ENSG00000133703 Transcript	ENST00000557334.5 protein_coding	2/3	231	34	12	G/R	GGT/CGT	rs121913530, CM076251, COSV55497461, COSV55497469, COSV55497582, COSV56157736

ESEfinder

Se observan 4 resultados con puntuaciones positivas en WT:

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	3.32010
211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	1.32130
216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	1.20390
220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	4.48380

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	-18.66980
211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	-10.04130
216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	-32.95400
220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	-0.89750

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	3.09350
211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	1.20920
216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	1.06100
220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	4.30110

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	-20.92360
211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	-11.43830
216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	-31.65020
220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	-0.94060

Se comparan las puntuaciones con las de la secuencia mutante:

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGAT	2.97960
211 (-272)	TGGTAGTTGGAGCTGATGGCGTAGGCAAGA	-9.66120
216 (-267)	GTTGGAGCTGATGGCGTAGGCAAGAGTGCC	1.43200
220 (-263)	GAGCTGATGGCGTAGGCAAGAGTGCCTTGA	4.32730

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGAT	-18.74990
211 (-272)	TGGTAGTTGGAGCTGATGGCGTAGGCAAGA	-11.37130
216 (-267)	GTTGGAGCTGATGGCGTAGGCAAGAGTGCC	-32.80890
220 (-263)	GAGCTGATGGCGTAGGCAAGAGTGCCTTGA	-1.39130

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGAT	2.62380
211 (-272)	TGGTAGTTGGAGCTGATGGCGTAGGCAAGA	-9.91910
216 (-267)	GTTGGAGCTGATGGCGTAGGCAAGAGTGCC	1.32740
220 (-263)	GAGCTGATGGCGTAGGCAAGAGTGCCTTGA	4.14180

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGAT	-20.94550
211 (-272)	TGGTAGTTGGAGCTGATGGCGTAGGCAAGA	-12.68870
216 (-267)	GTTGGAGCTGATGGCGTAGGCAAGAGTGCC	-31.55810
220 (-263)	GAGCTGATGGCGTAGGCAAGAGTGCCTTGA	-1.62400

Lo más probable es que se esté perdiendo un sitio *donor*, lo que no tendrá mucho efecto en el *splicing.*

En cuanto a los ESE, se producen algunas alteraciones que pueden estar afectando al *splicing:*

Pos	Seq	Score	Pos	Seq	Score	Pos	Seq	Score	Pos	Seq	Score
219 (-264)	GGAGCTC	-4.98070	219 (-264)	GGAGCTC	-3.65153	219 (-264)	GGAGCTCG	0.36938	219 (-264)	GGAGCTC	-3.32797
220 (-263)	GAGCTCG	-2.05624	220 (-263)	GAGCTCG	-2.17882	220 (-263)	GAGCTCGT	-1.55618	220 (-263)	GAGCTCG	-1.60679
221 (-262)	AGCTCGT	0.83257	221 (-262)	AGCTCGT	0.67513	221 (-262)	AGCTCGTG	3.55344	221 (-262)	AGCTCGT	-4.44566
222 (-261)	GCTCGTG	-3.46816	222 (-261)	GCTCGTG	-2.83859	222 (-261)	GCTCGTGG	-1.10005	222 (-261)	GCTCGTG	-1.80432
223 (-260)	CTCGTGG	-0.69252	223 (-260)	CTCGTGG	1.00274	223 (-260)	CTCGTGGC	-5.17740	223 (-260)	CTCGTGG	1.29005
224 (-259)	TCGTGGC	-3.59471	224 (-259)	TCGTGGC	-2.05833	224 (-259)	TCGTGGCG	-2.62991	224 (-259)	TCGTGGC	-2.06622
225 (-258)	CGTGGCG	-2.98536	225 (-258)	CGTGGCG	-0.38008	225 (-258)	CGTGGCGT	-4.32178	225 (-258)	CGTGGCG	-0.48562

Pos	Seq	Score	Pos	Seq	Score	Pos	Seq	Score	Pos	Seq	Score
219 (-264)	GGAGCTG	-3.50084	219 (-264)	GGAGCTG	-2.66251	219 (-264)	GGAGCTGG	-1.30618	219 (-264)	GGAGCTG	-2.48533
220 (-263)	GAGCTGG	0.52149	220 (-263)	GAGCTGG	-0.53152	220 (-263)	GAGCTGGT	-2.47149	220 (-263)	GAGCTGG	-1.30117
221 (-262)	AGCTGGT	0.22846	221 (-262)	AGCTGGT	0.46720	221 (-262)	AGCTGGTG	2.65576	221 (-262)	AGCTGGT	-5.26324
222 (-261)	GCTGGTG	-5.38319	222 (-261)	GCTGGTG	-4.15254	222 (-261)	GCTGGTGG	-3.79819	222 (-261)	GCTGGTG	-3.52864
223 (-260)	CTGGTGG	-0.93875	223 (-260)	CTGGTGG	0.62148	223 (-260)	CTGGTGGC	-7.48497	223 (-260)	CTGGTGG	1.07681
224 (-259)	TGGTGGC	-2.31648	224 (-259)	TGGTGGC	-1.08007	224 (-259)	TGGTGGCG	-0.59383	224 (-259)	TGGTGGC	-2.88777
225 (-258)	GGTGGCG	-4.56267	225 (-258)	GGTGGCG	-3.27828	225 (-258)	GGTGGCGT	-2.29326	225 (-258)	GGTGGCG	-2.63326

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	0	4	3	33	510.6662	16	-23.7770	8	15	44	500.3453	49	58.2806	56	116	0.48
mut	0	3	2	29	455.4963	15	-22.5634	6	15	41	464.2102	47	56.8045	49	109	0.45

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

HOT-SKIP

```
acgatacacgtctgcagtcaactggaattttcatgattgaattttgtaaggtattttgaaataattttttcatataaaggt
gagtttgtattaaaaggtactggtggagtatttgatagtgtattaaccttatgtgtgacatgttctaatatagtcacatt
ttcattattttattataagGCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGA
GTGCCTTGACGATACAGCTAATTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGgtaaatcttgttttaata
tgcatattactggtgcaggaccattctttgatacagataaaggtttctctgaccattttcatgagtacttattacaagat
aattatgctgaaagttaagttatctgaaatgtaccttgggtttcaagttatatgtaaccattaatatgggaactttactt
tc
```