Ejemplo comparación de resultados predictores in sillico

Cambio de estudio KRAS c.35G>A (chr12:25245350 G/A, rs121913529 o NM_033360.4: c.35G>A)

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
---------------------------------
               pos 5'->3'  phase strand  confidence  5'      exon intron    3'
                    79       0     +        0.00       TCATATAAAG^GTGAGTTTGT
                    97       1     +        0.37       GTATTAAAAG^GTACTGGTGG
                   303       0     +        0.65       AACAATAGAG^GTAAATCTTG

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

Acceptor splice sites, direct strand
------------------------------------
               pos 5'->3'  phase strand  confidence  5'    intron exon      3'
                   180       1     +        0.18       TTATTATAAG^GCCTGCTGAA
                   338       1     +        0.19       ACTGGTGCAG^GACCATTCTT

Acceptor splice sites, complement strand
----------------------------------------
   No acceptor site predictions above threshold.

--------------------------------------------------------------------------------
```

```
Donor splice sites, direct strand
---------------------------------
               pos 5'->3'  phase strand  confidence  5'      exon intron    3'
                    79       0     +        0.00       TCATATAAAG^GTGAGTTTGT
                    97       1     +        0.37       GTATTAAAAG^GTACTGGTGG
                   303       0     +        0.71       AACAATAGAG^GTAAATCTTG

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

Acceptor splice sites, direct strand
------------------------------------
               pos 5'->3'  phase strand  confidence  5'    intron exon      3'
                   180       1     +        0.21       TTATTATAAG^GCCTGCTGAA
                   216       1     +        0.07       CTTGTGGTAG^TTGGAGCTGA
                   222       1     +        0.07       GTAGTTGGAG^CTGATGGCGT
                   338       1     +        0.18       ACTGGTGCAG^GACCATTCTT

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 o 42 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	gtgagttt
90	104	0.79	ttaaaag	gtactggt
296	310	0.99	aatagag	gtaaatct

Donor site predictions for mut :

Start	End	Score	Exon	Intron
72	86	0.99	tataaag	gtgagttt
90	104	0.79	ttaaaag	gtactggt
296	310	0.99	aatagag	gtaaatct

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
160	200	0.69	tttcattatttttattata	agcctgctgaaaatgactgaa
336	376	0.67	caggaccattctttgatac	agataaaggtttctctgaccat

Acceptor site predictions for mut :

Start	End	Score	Intron	Exon
160	200	0.69	tttcattatttttattata	agcctgctgaaaatgactgaa
336	376	0.67	caggaccattctttgatac	agataaaggtttctctgaccat

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
agctg(g/a)tggcg	ggtggc	gatggc	28394	64%

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	0.383333333333	60	0	0	-3.0837295
mut	12	59	aaatgactg		1.10723640726	0.407407407407	54	0	0	-2.4512154
mut	12	55	gactgaata		0.297230706651	0.4	50	0	0	-2.51757
mut	12	49	atataaact		-1.22794910316	0.409090909091	44	0	0	-2.732024
mut	12	28	agctgatgg		1.75725074145	0.391304347826	23	0	0	-0.23965694

Aparece un nuevo BP en la secuencia mutante, pero tiene puntuación negativa, por lo que no se tendrá en cuenta.

Variant Effect Predictor tool

ENST00000557334.5:c.35G>A	12:25245350-25245350	T	missense_variant	KRAS	ENSG00000133703	Transcript	ENST00000256078.10	protein_coding	2/6	225	35	12	G/D	GGT/GAT	rs121913529, COSV55497369, COSV55497419, COSV55497479
ENST00000557334.5:c.35G>A	12:25245350-25245350	T	missense_variant	KRAS	ENSG00000133703	Transcript	ENST00000311936.8	protein_coding	2/5	225	35	12	G/D	GGT/GAT	rs121913529, COSV55497369, COSV55497419, COSV55497479
ENST00000557334.5:c.35G>A	12:25245350-25245350	T	missense_variant	KRAS	ENSG00000133703	Transcript	ENST00000556131.1	protein_coding	2/3	212	35	12	G/D	GGT/GAT	rs121913529, COSV55497369, COSV55497419, COSV55497479
ENST00000557334.5:c.35G>A	12:25245350-25245350	T	missense_variant	KRAS	ENSG00000133703	Transcript	ENST00000557334.5	protein_coding	2/3	232	35	12	G/D	GGT/GAT	rs121913529, COSV55497369, COSV55497419, COSV55497479

ESEfinder

Se observan 4 resultados con puntuaciones positivas en WT:

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	3.32010	198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	-18.66980	198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	3.09350	198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTGGT	-20.92360
211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	1.32130	211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	-10.04130	211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	1.20920	211 (-272)	TGGTAGTTGGAGCTGGTGGCGTAGGCAAGA	-11.43830
216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	1.20390	216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	-32.95400	216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	1.06100	216 (-267)	GTTGGAGCTGGTGGCGTAGGCAAGAGTGCC	-31.65020
220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	4.48380	220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	-0.89750	220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	4.30110	220 (-263)	GAGCTGGTGGCGTAGGCAAGAGTGCCTTGA	-0.94060

Se comparan las puntuaciones con las de la secuencia mutante:

198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTCGT	3.39360	198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTCGT	-18.77210	198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTCGT	3.17450	198 (-285)	GAATATAAACTTGTGGTAGTTGGAGCTCGT	-21.04820
211 (-272)	TGGTAGTTGGAGCTCGTGGCGTAGGCAAGA	-3.54770	211 (-272)	TGGTAGTTGGAGCTCGTGGCGTAGGCAAGA	-20.32670	211 (-272)	TGGTAGTTGGAGCTCGTGGCGTAGGCAAGA	-3.61290	211 (-272)	TGGTAGTTGGAGCTCGTGGCGTAGGCAAGA	-21.62670
216 (-267)	GTTGGAGCTCGTGGCGTAGGCAAGAGTGCC	1.20460	216 (-267)	GTTGGAGCTCGTGGCGTAGGCAAGAGTGCC	-30.52140	216 (-267)	GTTGGAGCTCGTGGCGTAGGCAAGAGTGCC	1.01780	216 (-267)	GTTGGAGCTCGTGGCGTAGGCAAGAGTGCC	-29.08920
220 (-263)	GAGCTCGTGGCGTAGGCAAGAGTGCCTTGA	4.60100	220 (-263)	GAGCTCGTGGCGTAGGCAAGAGTGCCTTGA	0.44220	220 (-263)	GAGCTCGTGGCGTAGGCAAGAGTGCCTTGA	4.40170	220 (-263)	GAGCTCGTGGCGTAGGCAAGAGTGCCTTGA	0.39760

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:*

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

219 (-264)	GGAGCTG	-3.50084	219 (-264)	GGAGCTG	-2.66251	219 (-264)	GGAGCTGA	-1.75608	219 (-264)	GGAGCTG	-2.48533
220 (-263)	GAGCTGA	1.24828	220 (-263)	GAGCTGA	-0.09615	220 (-263)	GAGCTGAT	-1.17460	220 (-263)	GAGCTGA	-3.68291
221 (-262)	AGCTGAT	-2.34928	221 (-262)	AGCTGAT	-1.31245	221 (-262)	AGCTGATG	2.29427	221 (-262)	AGCTGAT	-5.83107
222 (-261)	GCTGATG	-7.30280	222 (-261)	GCTGATG	-5.85893	222 (-261)	GCTGATGG	-3.37293	222 (-261)	GCTGATG	-0.97408
223 (-260)	CTGATGG	1.96982	223 (-260)	CTGATGG	2.34123	223 (-260)	CTGATGGC	-7.48497	223 (-260)	CTGATGG	1.23211
224 (-259)	TGATGGC	-4.38026	224 (-259)	TGATGGC	-2.46382	224 (-259)	TGATGGCG	0.70306	224 (-259)	TGATGGC	-0.27897
225 (-258)	GATGGCG	-4.11545	225 (-258)	GATGGCG	-3.57650	225 (-258)	GATGGCGT	-2.65412	225 (-258)	GATGGCG	-4.07843
226 (-257)	ATGGCGT	-0.54667	226 (-257)	ATGGCGT	-0.49481	226 (-257)	ATGGCGTA	-2.31901	226 (-257)	ATGGCGT	-3.48910

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	31	476.2297	15	-23.0149	7	16	43	485.6831	50	58.8533	51	116	0.44

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

HOT-SKIP

acgatacacgtctgcagtcaactggaattttcatgattgaattttgtaaggtattttgaaataatttttcatataaaggt
gagtttgtattaaaaggtactggtggagtatttgatagtgtattaaccttatgtgtgacatgttctaatatagtcacatt
ttcattatttttattataagGCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGA
GTGCCTTGACGATACAGCTAATTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGgtaaatcttgttttaata
tgcatattactggtgcaggaccattctttgatacagataaaggtttctctgaccattttcatgagtacttattacaagat
aattatgctgaaagttaagttatctgaaatgtaccttgggtttcaagttatatgtaaccattaatatgggaactttactt
tc