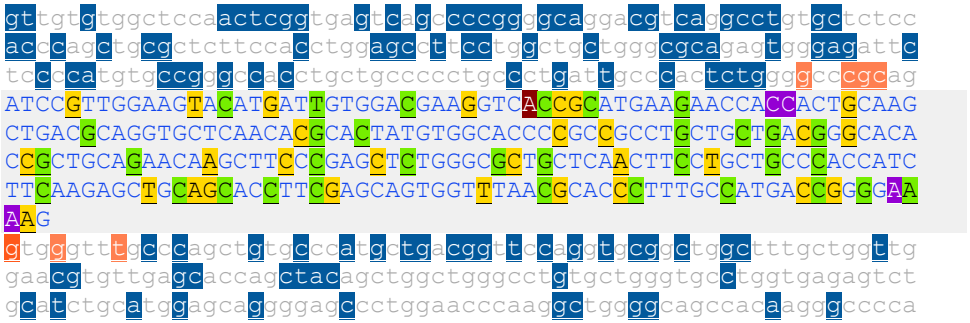
Ejemplo comparación de resultados predictores in sillico

Cambio de estudio SMARCA4 c.2843A>G (chr 19:11021759 A/G, COSV60793246 o NM_003072.5: c.2843A>G)

Exón 22 e intrones adyacentes:



El cambio se encuentra en segunda posición justo antes del exón 7 (la **a** 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'
                   20       -     +        0.00       CTCCAACTCG^GTGAGTCAGC
                  424       0     +        0.95       CGGGGAAAAG^GTGGGTTTGC H
                  462       2     +        0.36       ACGGTTCCAG^GTGCGGCTGG
                  534       1     +        0.00       TGGGTGCCTG^GTGAGAGTCT

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

Acceptor splice sites, direct strand
------------------------------------
              pos 5'->3'  phase strand  confidence  5'    intron exon     3'
                   66       1     +        0.00       CTCCACCCAG^CTGCGCTCTT
                  180       0     +        0.84       GGGCCCGCAG^ATCCGTTGGA
                  192       0     +        0.34       CCGTTGGAAG^TACATGATTG
                  211       1     +        0.17       GTGGACGAAG^GTCACCGCAT

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon     3'
       548          56       1     -        0.00       CTCCATGCAG^ATGCAGACTC
       402         202       0     -        0.18       CATGGCAAAG^GGTGCGTTAA
       378         226       0     -        0.17       CTGCTCGAAG^GTGCTGCAGC
       369         235       0     -        0.07       GGTGCTGCAG^CTCTTGAAGA
        99         505       0     -        0.17       CTGCGCCCAG^CAGCCAGGAA
```

```
Donor splice sites, direct strand
---------------------------------
              pos 5'->3'  phase strand  confidence  5'      exon intron    3'
                   20       -     +        0.00       CTCCAACTCG^GTGAGTCAGC
                  424       0     +        0.95       CGGGGAAAAG^GTGGGTTTGC H
                  462       2     +        0.36       ACGGTTCCAG^GTGCGGCTGG
                  534       1     +        0.00       TGGGTGCCTG^GTGAGAGTCT

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

Acceptor splice sites, direct strand
------------------------------------
              pos 5'->3'  phase strand  confidence  5'    intron exon     3'
                   66       1     +        0.00       CTCCACCCAG^CTGCGCTCTT
                  180       0     +        0.84       GGGCCCGCAG^ATCCGTTGGA
                  192       0     +        0.34       CCGTTGGAAG^TACATGATTG
                  211       1     +        0.14       GTGGACGAAG^GTCGCCGCAT

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon     3'
       548          56       1     -        0.00       CTCCATGCAG^ATGCAGACTC
       402         202       0     -        0.18       CATGGCAAAG^GGTGCGTTAA
       378         226       0     -        0.17       CTGCTCGAAG^GTGCTGCAGC
       369         235       0     -        0.07       GGTGCTGCAG^CTCTTGAAGA
        99         505       0     -        0.17       CTGCGCCCAG^CAGCCAGGAA
```

A pesar de que hay un cambio en uno de los sitios *acceptor* (en amarillo) por la presencia de la mutación, este no se tendrá en cuenta dado que no se emplea en el *splicing* y, al debilitarse, con menos probabilidad aún.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
13	27	0.99	caactcg	gt gagtca
417	431	0.86	ggaaaag	gt gggttt
455	469	0.91	gttccag	gt gcggct
527	541	0.87	gtgcctg	gt gagagt

Donor site predictions for mut :

Start	End	Score	Exon	Intron
13	27	0.99	caactcg	gt gagtca
417	431	0.86	ggaaaag	gt gggttt
455	469	0.91	gttccag	gt gcggct
527	541	0.87	gtgcctg	gt gagagt

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
46	86	0.64	aggcctgtgctctccaccc	ag ctgcgctcttccacctggag

Acceptor site predictions for mut :

Start	End	Score	Intron	Exon
46	86	0.64	aggcctgtgctctccaccc	ag ctgcgctcttccacctggag

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
aggtc(a/g)ccgca	aggtca	aggtcg	28621	65%

Human Splicing Finder

Alteration of auxiliary sequences	Significant alteration of ESE / ESS motifs ratio (-2)

Algorithm/Matix	position	sequence
ESE_9G8 (ESE Site Broken)	chr19:11021755	GGTCAC
ESE_9G8 (New ESE Site)	chr19:11021756	GTCGCC
PESE (New ESE Site)	chr19:11021756	GTCGCCGC
EIE (ESE Site Broken)	chr19:11021757	TCACCG
ESE_SRp40 (ESE Site Broken)	chr19:11021757	TCACCGC
EIE (ESE Site Broken)	chr19:11021759	ACCGCA

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr		
wt	24	51	acatgattg		-1.34220291425	0.434782608696	46	0	0	-2.8950587	
wt	24	35	aggtcaccg		0.361138500933	0.466666666667	30	0	0	-1.2050493	
wt	24	27	gcatgaaga		-2.07477000634	0.454545454545	22	0	0	-1.6563508	
mut	24	51	acatgattg		-1.34220291425	0.434782608696	46	0	0	-2.8950587	
mut	24	27	gcatgaaga		-2.07477000634	0.454545454545	22	0	0	-1.6563508	

Desaparece un BP en la secuencia mutante, pero, como tiene puntuación negativa, no lo tendremos en cuenta.

Variant Effect Predictor tool

ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000344626.10	protein_coding	19/35	2827	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000413806.7	protein_coding	22/37	2843	2843	948	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000429416.8	protein_coding	20/36	2932	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000444061.8	protein_coding	20/35	2747	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000450717.7	protein_coding	18/35	2651	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000541122.6	protein_coding	20/35	2936	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	non_coding_transcript_exon_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000585799.5	retained_intron	1/16	1179	-	-	-	-	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	non_coding_transcript_exon_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000586122.5	processed_transcript	4/5	233	-	-	-	-	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	non_coding_transcript_exon_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000587988.5	retained_intron	3/4	300	-	-	-	-	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000589677.5	protein_coding	20/35	2981	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000590574.6	protein_coding	19/34	3088	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000591545.6	protein_coding	18/24	2524	2525	842	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	non_coding_transcript_exon_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000591595.5	retained_intron	2/17	723	-	-	-	-	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	non_coding_transcript_exon_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000592604.6	retained_intron	4/19	982	-	-	-	-	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000642350.1	protein_coding	11/27	1136	1136	379	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000642508.1	protein_coding	1/18	12	14	5	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000642628.1	protein_coding	19/35	2900	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000642726.1	protein_coding	19/35	2922	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000643208.1	protein_coding	11/27	1208	1208	403	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000643296.1	protein_coding	19/34	2900	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000643534.1	protein_coding	10/25	936	938	313	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000643549.1	protein_coding	19/35	2922	2651	884	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000643857.1	protein_coding	10/25	1006	1007	336	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000643995.1	protein_coding	16/32	2065	2066	689	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	missense_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000644065.1	protein_coding	12/27	1378	1379	460	H/R	CAC/CGC	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	non_coding_transcript_exon_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000644267.1	retained_intron	10/21	1006	-	-	-	-	COSV60793246, COSV60804891
ENST00000413806.7:c.2843A>G	19:11021759-11021759	G	3_prime_UTR_variant, NMD_transcript_variant	SMARCA4	ENSG00000127616	Transcript	ENST00000644290.1	nonsense_mediated_decay	10/16	908	-	-	-	-	COSV60793246, COSV60804891

ESEfinder

Se encuentran 2 resultados con puntuaciones positivas:

197 (-407)	TGATTGTGGACGAAGGTCACCGCATGAAGA	1.81880	197 (-407)	TGATTGTGGACGAAGGTCACCGCATGAAGA	0.25640	197 (-407)	TGATTGTGGACGAAGGTCACCGCATGAAGA	1.80450	197 (-407)	TGATTGTGGACGAAGGTCACCGCATGAAGA	0.05100
211 (-393)	GGTCACCGCATGAAGAACCACCACTGCAAG	-26.07770	211 (-393)	GGTCACCGCATGAAGAACCACCACTGCAAG	0.34490	211 (-393)	GGTCACCGCATGAAGAACCACCACTGCAAG	-22.33680	211 (-393)	GGTCACCGCATGAAGAACCACCACTGCAAG	0.24590

Comparando con la secuencia mutante:

197 (-407)	TGATTGTGGACGAAGGTCGCCGCATGAAGA	-0.48380	197 (-407)	TGATTGTGGACGAAGGTCGCCGCATGAAGA	0.47350	197 (-407)	TGATTGTGGACGAAGGTCGCCGCATGAAGA	-0.59070	197 (-407)	TGATTGTGGACGAAGGTCGCCGCATGAAGA	0.21090
211 (-393)	GGTCGCCGCATGAAGAACCACCACTGCAAG	-26.13430	211 (-393)	GGTCGCCGCATGAAGAACCACCACTGCAAG	0.96490	211 (-393)	GGTCGCCGCATGAAGAACCACCACTGCAAG	-22.40360	211 (-393)	GGTCGCCGCATGAAGAACCACCACTGCAAG	0.98490

Por lo tanto, se estaría fortaleciendo un sitio *acceptor* y perdiendo un *donor*.

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

209 (-395)	AAGGTCA	-4.17237	209 (-395)	AAGGTCA	-3.29495	209 (-395)	AAGGTCAC	-4.70950	209 (-395)	AAGGTCA	-4.25813
210 (-394)	AGGTCAC	-3.84479	210 (-394)	AGGTCAC	-2.68249	210 (-394)	AGGTCACC	-2.02023	210 (-394)	AGGTCAC	-3.68763
211 (-393)	GGTCACC	-6.04711	211 (-393)	GGTCACC	-4.65974	211 (-393)	GGTCACCG	4.76587	211 (-393)	GGTCACC	0.80299
212 (-392)	GTCACCG	0.58474	212 (-392)	GTCACCG	-0.01121	212 (-392)	GTCACCGC	-1.47433	212 (-392)	GTCACCG	-2.49916
213 (-391)	TCACCGC	-3.59706	213 (-391)	TCACCGC	-1.97940	213 (-391)	TCACCGCA	-0.65439	213 (-391)	TCACCGC	3.74479
214 (-390)	CACCGCA	2.41369	214 (-390)	CACCGCA	3.13987	214 (-390)	CACCGCAT	-0.11867	214 (-390)	CACCGCA	-3.93648
215 (-389)	ACCGCAT	-3.48111	215 (-389)	ACCGCAT	-2.14197	215 (-389)	ACCGCATG	-1.31143	215 (-389)	ACCGCAT	-3.53163

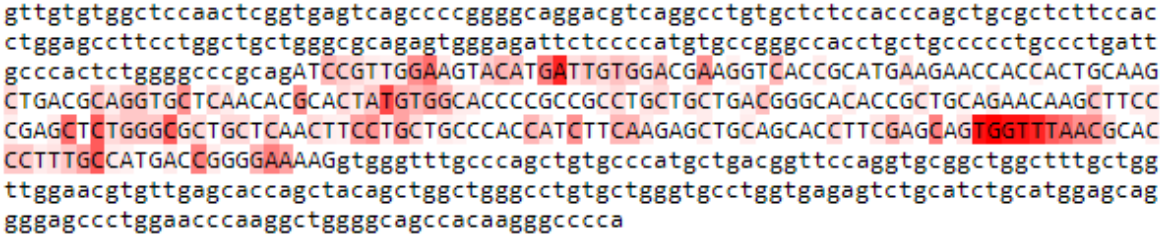
209 (-395)	AAGGTCG	-4.89915	209 (-395)	AAGGTCG	-3.73032	209 (-395)	AAGGTCGC	-6.00639	209 (-395)	AAGGTCG	-1.87640
210 (-394)	AGGTCGC	-1.26706	210 (-394)	AGGTCGC	-0.90283	210 (-394)	AGGTCGCC	-1.65875	210 (-394)	AGGTCGC	-3.11980
211 (-393)	GGTCGCC	-4.12749	211 (-393)	GGTCGCC	-2.95336	211 (-393)	GGTCGCCG	4.34061	211 (-393)	GGTCGCC	-1.75158
212 (-392)	GTCGCCG	-2.32384	212 (-392)	GTCGCCG	-1.73097	212 (-392)	GTCGCCGC	-1.47433	212 (-392)	GTCGCCG	-2.65446
213 (-391)	TCGCCGC	-1.53327	213 (-391)	TCGCCGC	-0.59565	213 (-391)	TCGCCGCA	-1.95128	213 (-391)	TCGCCGC	1.13599
214 (-390)	CGCCGCA	1.96647	214 (-390)	CGCCGCA	3.43810	214 (-390)	CGCCGCAT	0.24219	214 (-390)	CGCCGCA	-2.49131
215 (-389)	GCCGCAT	-2.55323	215 (-389)	GCCGCAT	-1.90442	215 (-389)	GCCGCATG	0.44064	215 (-389)	GCCGCAT	-4.98635

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	5	2	30	430.8032	10	-14.0759	13	24	90	1066.8624	99	132.6552	47	226	0.21
mut	0	5	2	30	430.8032	10	-14.2021	13	24	88	1052.2509	99	132.6487	47	224	0.21

Both alleles have a comparable chance of exon skipping.

HOT-SKIP


gttgtgtggctccaactcggtgagtcagccccggggcaggacgtcaggcctgtgctctccacccagctgcgctcttccac
ctggagccttcctggctgctgggcgcagagtgggagattctccccatgtgccgggccacctgctgccccctgccctgatt
gcccactctggggcccgcagATCCGTTGGAAGTACATGATTGTGGACGAAGGTCACCGCATGAAGAACCACCACTGCAAG
CTGACGCAGGTGCTCAACACGCACTATGTGGCACCCCGCCGCCTGCTGCTGACGGGCACACCGCTGCAGAACAAGCTTCC
CGAGCTCTGGGCGCTGCTCAACTTCCTGCTGCCCACCATCTTCAAGAGCTGCAGCACCTTCGAGCAGTGGTTTAACGCAC
CCTTTGCCATGACCGGGGAAAAGgtgggtttgcccagctgtgcccatgctgacggttccaggtgcggctggctttgctgg
ttggaacgtgttgagcaccagctacagctggctgggcctgtgctgggtgcctggtgagagtctgcatctgcatggagcag
gggagccctggaacccaaggctggggcagccacaagggcccca