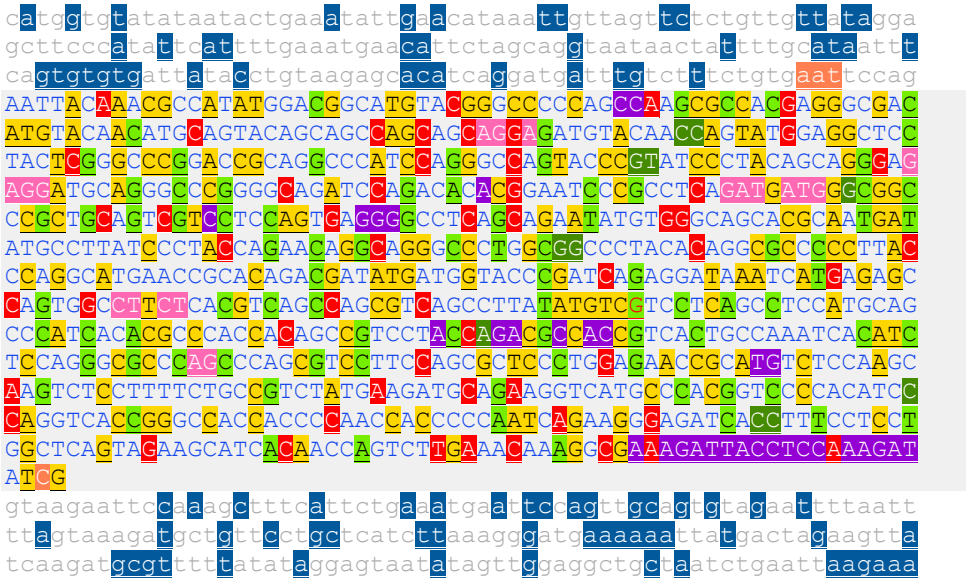
Ejemplo comparación de resultados predictores in sillico

Cambio de estudio ARID1B c.2262G>A (chr6:157201166 G/A, rs11547292 o NM_001374828.1: c.2262G>A)

Exón 15 e intrones adyacentes:



El cambio se encuentra en la octava línea del exón 15 (la **g** en color rojo subrayada en verde).

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'
98	1	+	0.41	ATTCTAGCAG^GTAATAACTA
844	0	+	0.32	CACATCCCAG^GTCACCGGGC
965	1	+	0.67	AAAGATATCG^GTAAGAATTC

Donor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' exon intron 3'
690	455	2	-	0.70	GTGGCGTCTG^GTAGGACGGC
540	605	0	-	0.76	TCATGCCTGG^GTAAGGGGGC

Acceptor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' intron exon 3'
58	1	+	0.00	GTTGTTATAG^GAGCTTCCCA
143	2	+	0.18	TACCTGTAAG^AGCACATCAG
145	1	+	0.19	CCTGTAAGAG^CACATCAGGA
153	0	+	0.25	AGCACATCAG^GATGATTTGT
180	0	+	0.96	TGAATTCCAG^AATTACAAAC H
498	0	+	0.14	TCCCTACCAG^AACAGGCAGG
503	2	+	0.25	ACCAGAACAG^GCAGGGCCCT
544	1	+	0.26	CCTTACCCAG^GCATGAACCG
649	1	+	0.17	TCGTCCTCAG^CCTCCATGCA
660	0	+	0.19	CTCCATGCAG^CCCATCACAC
734	2	+	0.07	GGGCGCCCAG^CCCAGCGTCC
739	1	+	0.18	CCCAGCCCAG^CGTCCTTCCA
750	0	+	0.19	GTCCTTCCAG^CGCTCCCTGG
762	0	+	0.19	CTCCCTGGAG^AACCGCATGT
779	2	+	0.17	TGTCTCCAAG^CAAGTCTCCT
783	0	+	0.07	TCCAAGCAAG^TCTCCTTTTC

Acceptor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' intron exon 3'
800	345	1	-	0.85	ATCTTCATAG^ACGGCAGAAA
793	352	2	-	0.17	TAGACGGCAG^AAAAGGAGAC
788	357	1	-	0.07	GGCAGAAAAG^GAGACTTGCT
348	797	1	-	0.95	CCTGCTGTAG^GGATACGGGT
137	1008	0	-	0.28	GCTCTTACAG^GTATAATCAC

Donor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' exon intron 3'
98	1	+	0.41	ATTCTAGCAG^GTAATAACTA
844	0	+	0.32	CACATCCCAG^GTCACCGGGC
965	1	+	0.67	AAAGATATCG^GTAAGAATTC

Donor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' exon intron 3'
690	455	2	-	0.71	GTGGCGTCTG^GTAGGACGGC
540	605	0	-	0.72	TCATGCCTGG^GTAAGGGGGC

Acceptor splice sites, direct strand

pos 5'->3'	phase	strand	confidence	5' intron exon 3'
58	1	+	0.00	GTTGTTATAG^GAGCTTCCCA
143	2	+	0.18	TACCTGTAAG^AGCACATCAG
145	1	+	0.19	CCTGTAAGAG^CACATCAGGA
153	0	+	0.25	AGCACATCAG^GATGATTTGT
180	0	+	0.96	TGAATTCCAG^AATTACAAAC H
498	0	+	0.14	TCCCTACCAG^AACAGGCAGG
503	2	+	0.25	ACCAGAACAG^GCAGGGCCCT
544	1	+	0.27	CCTTACCCAG^GCATGAACCG
649	1	+	0.18	TCATCCTCAG^CCTCCATGCA
660	0	+	0.25	CTCCATGCAG^CCCATCACAC
734	2	+	0.07	GGGCGCCCAG^CCCAGCGTCC
739	1	+	0.18	CCCAGCCCAG^CGTCCTTCCA
750	0	+	0.19	GTCCTTCCAG^CGCTCCCTGG
762	0	+	0.19	CTCCCTGGAG^AACCGCATGT
779	2	+	0.17	TGTCTCCAAG^CAAGTCTCCT
783	0	+	0.07	TCCAAGCAAG^TCTCCTTTTC

Acceptor splice sites, complement strand

pos 3'->5'	pos 5'->3'	phase	strand	confidence	5' intron exon 3'
800	345	1	-	0.85	ATCTTCATAG^ACGGCAGAAA
793	352	2	-	0.17	TAGACGGCAG^AAAAGGAGAC
788	357	1	-	0.07	GGCAGAAAAG^GAGACTTGCT
348	797	1	-	0.95	CCTGCTGTAG^GGATACGGGT
137	1008	0	-	0.28	GCTCTTACAG^GTATAATCAC

Se produce una alteración en uno de los sitios *acceptor* (en amarillo) de la secuencia WT a la mutante debido a la presencia de la mutación. Este cambio hace que este sitio *acceptor* tenga un *score* 0,01 mayor. Apesar de que la variación no es muy grande, en el caso de que fuera suficiente para que el *spliceosome* reconociera este sitio y no el sitio *acceptor* normal para el exón, se produciría la pérdida de los primeros 469 nucleótidos del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for wt :

Start	End	Score	Exon	Intron
91	105	0.78	ctagcag	**gt**aataac
132	146	0.82	tatacct	**gt**aagagc
958	972	0.99	gatatcg	**gt**aagaat

Donor site predictions for mut :

Start	End	Score	Exon	Intron
91	105	0.78	ctagcag	**gt**aataac
132	146	0.82	tatacct	**gt**aagagc
958	972	0.99	gatatcg	**gt**aagaat

Acceptor site predictions for wt :

Start	End	Score	Intron	Exon
38	78	0.86	gttagttctctgttgttat	**ag**gagcttcccatattcatttt
333	373	0.67	ccagtacccgtatccctac	**ag**cagggagaggatgcagggcc
508	548	0.42	ggccctggcggccctacac	**ag**gcgccccttacccaggcat
524	564	0.73	cacaggcgccccttaccc	**ag**gcatgaaccgcacagacgat
823	863	0.71	cccacggtccccacatccc	**ag**gtcaccgggccaccacccca
1037	1077	0.72	ctgttcctgctcatcttaa	**ag**ggatgaaaaaattatgacta
1085	1125	0.78	tcaagatgcgtttttatat	**ag**gagtaatatagttggaggct

Acceptor site predictions for mut :

Start	End	Score	Intron	Exon
38	78	0.86	gttagttctctgttgttat	**ag**gagcttcccatattcatttt
333	373	0.67	ccagtacccgtatccctac	**ag**cagggagaggatgcagggcc
508	548	0.42	ggccctggcggccctacac	**ag**gcgccccttacccaggcat
524	564	0.73	cacaggcgccccttaccc	**ag**gcatgaaccgcacagacgat
823	863	0.71	cccacggtccccacatccc	**ag**gtcaccgggccaccacccca
1037	1077	0.72	ctgttcctgctcatcttaa	**ag**ggatgaaaaaattatgacta
1085	1125	0.78	tcaagatgcgtttttatat	**ag**gagtaatatagttggaggct

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
atgtc(g/a)tcctc	tgtcgt	tgtcat	29829	73%

Human Splicing Finder

Alteration of auxiliary sequences	Significant alteration of ESE / ESS motifs ratio (-3)

Algorithm/Matix	position	sequence
PESS (New ESS Site)	chr6:157201159	ATATGTCA
ESE_9G8 (ESE Site Broken)	chr6:157201163	GTCGTC
ESE_SC35 (ESE Site Broken)	chr6:157201166	GTCCTCAG

New Donor splice site	Activation of a cryptic Donor site. Potential alteration of splicing

Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr6:157201160	- REF : TATGTCGTC - ALT : TATGTCATC	61.93 > 72.23 => 16.63%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	29	64	tcatgagag	-1.96618879083	0.610169491525	11	9	17	-0.70892329
wt	29	46	ttctcacgt	2.2032093763	0.634146341463	26	15	22	0.028441393
wt	29	41	acgtcagcc	-1.93930160981	0.638888888889	21	15	22	-1.2755241
wt	29	31	gcgtcagcc	-1.20509861024	0.692307692308	11	15	22	-0.33781305
wt	29	25	gccttatat	-1.41938724254	0.7	5	15	22	-0.039442858
mut	29	64	tcatgagag	-1.96618879083	0.610169491525	11	9	17	-0.70892329
mut	29	46	ttctcacgt	2.2032093763	0.634146341463	26	15	22	0.028441393
mut	29	41	acgtcagcc	-1.93930160981	0.638888888889	21	15	22	-1.2755241
mut	29	31	gcgtcagcc	-1.20509861024	0.692307692308	11	15	22	-0.33781305
mut	29	25	gccttatat	-1.41938724254	0.7	5	15	22	-0.039442858
mut	29	18	atgtcatcc	-1.56492676376	0.769230769231	1	12	19	0.15117681

Se predice un nuevo sitio BP en la secuencia mutante, lo que es probable que afecte al *splicing*.

Variant Effect Predictor tool

ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000346085.10	protein_coding	19/21	5747	4821	1607	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000350026.10	protein_coding	17/19	4533	4533	1511	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000414678.7	protein_coding	17/19	3099	3099	1033	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000635849.1	protein_coding	15/17	2849	2262	754	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000635957.1	protein_coding	11/13	1893	1893	631	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	non_coding_transcript_exon_variant	ARID1B	ENSG00000049618	Transcript	ENST00000636227.1	retained_intron	3/5	3404	-	-	-	-	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	non_coding_transcript_exon_variant	ARID1B	ENSG00000049618	Transcript	ENST00000636254.1	retained_intron	2/4	861	-	-	-	-	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000636930.2	protein_coding	18/20	5244	4941	1647	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	non_coding_transcript_exon_variant	ARID1B	ENSG00000049618	Transcript	ENST00000636940.1	retained_intron	9/11	2938	-	-	-	-	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000637015.1	protein_coding	12/14	2309	2310	770	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	3_prime_UTR_variant, NMD_transcript_variant	ARID1B	ENSG00000049618	Transcript	ENST00000637568.1	nonsense_mediated_decay	13/15	2223	-	-	-	-	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	non_coding_transcript_exon_variant	ARID1B	ENSG00000049618	Transcript	ENST00000637741.1	processed_transcript	9/11	1607	-	-	-	-	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000637810.1	protein_coding	13/15	2425	2283	761	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000637904.1	protein_coding	16/18	2943	2442	814	S	TCG/TCA	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	upstream_gene_variant	ARID1B	ENSG00000049618	Transcript	ENST00000637933.1	retained_intron	-	-	-	-	-	-	rs368163089
ENST00000635849.1:c.2262G>A	6:157201166-157201166	A	synonymous_variant	ARID1B	ENSG00000049618	Transcript	ENST00000647938.1	protein_coding	18/20	4572	4572	1524	S	TCG/TCA	rs368163089

ESEfinder

Solo se observa un resultado con puntuación positiva para las matrices 3', pero cuando se comparan con la secuencia mutante la variación descendente es muy limitada por lo que el sitio sector más débil es poco probable:

635 (-510)	ATATGTCGTCCTCAGCCTCCATGCAGCCCA	-18.50440	635 (-510)	ATATGTCGTCCTCAGCCTCCATGCAGCCCA	5.70930	635 (-510)	ATATGTCGTCCTCAGCCTCCATGCAGCCCA	-18.87630	635 (-510)	ATATGTCGTCCTCAGCCTCCATGCAGCCCA	5.37770
635 (-510)	ATATGTCATCCTCAGCCTCCATGCAGCCCA	-18.46360	635 (-510)	ATATGTCATCCTCAGCCTCCATGCAGCCCA	5.41780	635 (-510)	ATATGTCATCCTCAGCCTCCATGCAGCCCA	-18.79890	635 (-510)	ATATGTCATCCTCAGCCTCCATGCAGCCCA	5.02830

En cuando a las ESE, hay algunos cambios entre las secuencias (WT arriba, mutante abajo) que podrían estar afectando al *splicing:*

635 (-510)	ATATGTC	-6.73029	635 (-510)	ATATGTC	-4.76730	635 (-510)	ATATGTCG	-0.46075	635 (-510)	ATATGTC	-2.84165
636 (-509)	TATGTCG	-7.40824	636 (-509)	TATGTCG	-5.38722	636 (-509)	TATGTCGT	-4.56532	636 (-509)	TATGTCG	0.73474
637 (-508)	ATGTCGT	-0.08896	637 (-508)	ATGTCGT	-0.43562	637 (-508)	ATGTCGTC	-1.66198	637 (-508)	ATGTCGT	-4.14941
638 (-507)	TGTCGTC	-5.04297	638 (-507)	TGTCGTC	-3.05621	638 (-507)	TGTCGTCC	-1.69822	638 (-507)	TGTCGTC	-0.96416
639 (-506)	GTCGTCC	-6.32742	639 (-506)	GTCGTCC	-4.53178	639 (-506)	GTCGTCCT	-0.55801	639 (-506)	GTCGTCC	-2.00585
640 (-505)	TCGTCCT	-3.71493	640 (-505)	TCGTCCT	-2.30100	640 (-505)	TCGTCCTC	-2.43265	640 (-505)	TCGTCCT	-3.09336
641 (-504)	CGTCCTC	-1.48837	641 (-504)	CGTCCTC	0.25677	641 (-504)	CGTCCTCA	1.03086	641 (-504)	CGTCCTC	-0.50328

635 (-510)	ATATGTC	-6.73029	635 (-510)	ATATGTC	-4.76730	635 (-510)	ATATGTCA	-0.91065	635 (-510)	ATATGTC	-2.84165
636 (-509)	TATGTCA	-6.68146	636 (-509)	TATGTCA	-4.95185	636 (-509)	TATGTCAT	-3.26843	636 (-509)	TATGTCA	-1.64699
637 (-508)	ATGTCAT	-2.66670	637 (-508)	ATGTCAT	-2.21527	637 (-508)	ATGTCATC	-2.02347	637 (-508)	ATGTCAT	-4.71724
638 (-507)	TGTCATC	-6.96258	638 (-507)	TGTCATC	-4.76260	638 (-507)	TGTCATCC	-1.27295	638 (-507)	TGTCATC	1.59040
639 (-506)	GTCATCC	-3.41885	639 (-506)	GTCATCC	-2.81203	639 (-506)	GTCATCCT	-0.55801	639 (-506)	GTCATCC	-1.85055
640 (-505)	TCATCCT	-5.77872	640 (-505)	TCATCCT	-3.68476	640 (-505)	TCATCCTC	-1.13576	640 (-505)	TCATCCT	-0.48456
641 (-504)	CATCCTC	-1.04114	641 (-504)	CATCCTC	-0.04146	641 (-504)	CATCCTCA	0.67001	641 (-504)	CATCCTC	-1.94844

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	8	4	3	58	694.8816	29	-49.5517	30	60	261	3127.9790	228	354.1608	102	579	0.18
mut	9	4	3	58	699.6026	31	-51.2153	30	60	261	3127.9790	229	354.9427	105	580	0.18

Both alleles have a comparable chance of exon skipping.

HOT-SKIP

catggtgtatataatactgaaatattgaacataaattgttagttctctgttgttataggagcttcccatattcattttga
aatgaacattctagcaggtaataactattttgcataatttcagtgtgtgattatacctgtaagagcacatcaggatgatt
tgtctttctgtgaattccagAATTACAAACGCCATATGGACGGCATGTACGGGCCCCCAGCCAAGCGCCACGAGGGCGAC
ATGTACAACATGCAGTACAGCAGCCAGCAGCAGGAGATGTACAACCAGTATGGAGGCTCCTACTCGGGCCCGGACCGCAG
GCCCATCCAGGGCCAGTACCCGTATCCCTACAGCAGGGAGAGGATGCAGGGCCCGGGGCAGATCCAGACACACGGAATCC
CGCCTCAGATGATGGGCGGCCCGCTGCAGTCGTCCTCCAGTGAGGGGCCTCAGCAGAATATGTGGGCAGCACGCAATGAT
ATGCCTTATCCCTACCAGAACAGGCAGGGCCCTGGCGGCCCTACACAGGCGCCCCCTTACCCAGGCATGAACCGCACAGA
CGATATGATGGTACCCGATCAGAGGATAAATCATGAGAGCCAGTGGCCTTCTCACGTCAGCCAGCGTCAGCCTTATATGT
CGTCCTCAGCCTCCATGCAGCCCATCACACGCCCACCACAGCCGTCCTACCAGACGCCACCGTCACTGCCAAATCACATC
TCCAGGGCGCCCAGCCCAGCGTCCTTCCAGCGCTCCCTGGAGAACCGCATGTCTCCAAGCAAGTCTCCTTTTCTGCCGTC
TATGAAGATGCAGAAGGTCATGCCCACGGTCCCCACATCCCAGGTCACCGGGCCACCACCCCAACCACCCCCAATCAGAA
GGGAGATCACCTTTCCTCCTGGCTCAGTAGAAGCATCACAACCAGTCTTGAAACAAAGGCGAAAGATTACCTCCAAAGAT
ATCGgtaagaattccaaagctttcattctgaaatgaattccagttgcagtgtagaattttaattttagtaaagatgctgt
tcctgctcatcttaaagggatgaaaaaattatgactagaagttatcaagatgcgtttttatataggagtaatatagttgg
aggctgctaatctgaattaagaaa

1841	462	G	ATATGTCGTCCTCAG	ATGTCGTCCTC	0	0	0	1	9.6909	0	-0.3363	0	0	0	0.0000	0	1.1950	1	0	1.00
1842	462	A	ATATGTCATCCTCAG	ATGTCATCCTC	1	0	0	1	14.4119	2	-1.9999	0	0	0	0.0000	1	1.9769	4	1	4.00
1843	462	C	ATATGTCCTCCTCAG	ATGTCCTCCTC	0	0	0	1	15.8370	0	-0.2448	0	0	0	0.0000	3	4.1556	1	3	0.33
1844	462	T	ATATGTCTTCCTCAG	ATGTCTTCCTC	2	0	0	4	54.2818	2	-2.3363	1	0	0	0.0000	1	1.3094	8	2	4.00