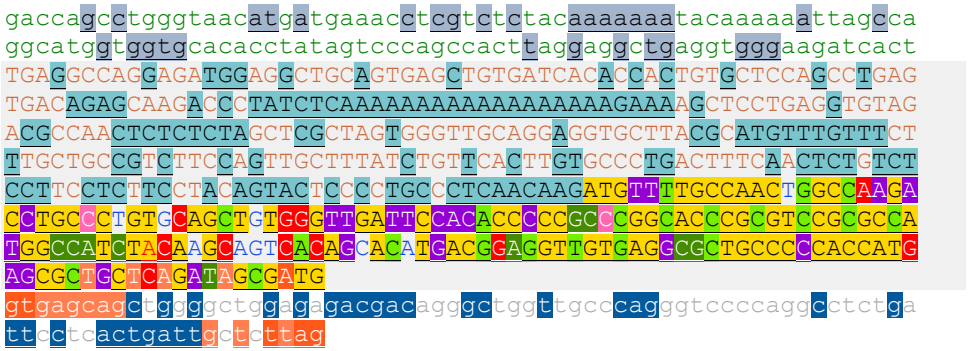
Ejemplo comparación de resultados predictores in sillico

Cambio de estudio TP53 c.92A>G (chr 17:7675124 A/G, rs148924904 o NM_000546.6: c.92A>G)

Exón 1 e intrones adyacentes:



El cambio se encuentra en la penúltima fila del exón 1 (la **a** 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'
                  562       1     +        0.00       GATAGCGATG^GTGAGCAGCT

Donor splice sites, complement strand
-------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'      exon intron    3'
       286         357       1     -        0.59       ACAAACATGC^GTAAGCACCT

Acceptor splice sites, direct strand
------------------------------------
              pos 5'->3'  phase strand  confidence  5'    intron exon      3'
                  275       0     +        0.15       TGGGTTGCAG^GAGGTGCTTA
                  317       0     +        0.18       CGTCTTCCAG^TTGCTTTATC
                  377       0     +        0.97       CTTCCTACAG^TACTCCCCTG
                  398       0     +        0.07       CCTCAACAAG^ATGTTTTGCC

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'
                  562       1     +        0.00       GATAGCGATG^GTGAGCAGCT

Donor splice sites, complement strand
-------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'      exon intron    3'
       286         357       1     -        0.59       ACAAACATGC^GTAAGCACCT

Acceptor splice sites, direct strand
------------------------------------
              pos 5'->3'  phase strand  confidence  5'    intron exon      3'
                  275       0     +        0.15       TGGGTTGCAG^GAGGTGCTTA
                  317       0     +        0.18       CGTCTTCCAG^TTGCTTTATC
                  377       0     +        0.97       CTTCCTACAG^TACTCCCCTG

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon      3'
       548          95       1     -        0.07       GCTATCTGAG^CAGCGCTCAT
       545          98       1     -        0.14       ATCTGAGCAG^CGCTCATGGT
       527         116       0     -        0.18       GTGGGGGCAG^CGCCTCACAA
       488         155       0     -        0.33       CTGCTTGCAG^ATGGCCATGG
```

El cambio que existe entre la predicción para la secuencia *wild type* y la secuencia mutante es la desaparición de uno de los sitios *acceptor* en la secuencia mutante, el cual se encuentra dentro del exón y ni siquiera está cerca de la posición de interés por lo que es probable que desaparecido porque su confianza es muy baja y no se tendrá en cuenta.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.0.139.572061.0 :

Start	End	Score	Exon	Intron
555	569	0.93	agcgatg	**gt**gagcag

Donor site predictions for 10.42.1.119.572073.0 :

Start	End	Score	Exon	Intron
555	569	0.93	agcgatg	**gt**gagcag

Acceptor site predictions for 10.42.0.139.572061.0 :

Start	End	Score	Intron	Exon
237	277	0.79	gtagacgccaactctctct	**ag**ctcgctagtgggttgcagga
297	337	0.69	ttctttgctgccgtcttcc	**ag**ttgctttatctgttcacttg
357	397	0.99	gtctccttcctcttcctac	**ag**tactcccctgccctcaacaa

Acceptor site predictions for 10.42.1.119.572073.0 :

Start	End	Score	Intron	Exon
237	277	0.79	gtagacgccaactctctct	**ag**ctcgctagtgggttgcagga
297	337	0.69	ttctttgctgccgtcttcc	**ag**ttgctttatctgttcacttg
357	397	0.99	gtctccttcctcttcctac	**ag**tactcccctgccctcaacaa

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
catct(a/g)caagc	tctaca	tctgca	28412	64%

CRYP-SKIP

Parece que hay un sitio críptico de *splicing* dentro del propio exón, pero el cambio de interés (la primera **a** en minúsculas detrás de las mayúsculas, que indican el exón) no lo toma en consideración, por lo que no debe considerar que tenga algún efecto en el *splicing*.

Human Splicing Finder

Alteration of auxiliary sequences	Significant alteration of ESE / ESS motifs ratio (-4)

Algorithm/Matix	position	sequence
RESCUE ESE (ESE Site Broken)	chr17:7675125	TACAAG
EIE (ESE Site Broken)	chr17:7675126	CTACAA
ESE_SRp40 (ESE Site Broken)	chr17:7675126	CTACAAG
Sironi_motif1 (ESS Site Broken)	chr17:7675126	CTACAAGC
EIE (ESE Site Broken)	chr17:7675127	TCTACA
PESE (New ESE Site)	chr17:7675127	TCTGCAAG
EIE (ESE Site Broken)	chr17:7675128	ATCTAC
EIE (ESE Site Broken)	chr17:7675129	CATCTA

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	12	82	cagtcacag	-0.291165564125	0.441558441558	27	11	15	-1.1389338
wt	12	72	acatgacgg	0.390478992917	0.44776119403	17	11	15	-0.23705118
wt	12	60	ttgtgaggc	-1.03767995916	0.472727272727	5	11	15	-0.028599734
wt	12	41	ccatgagcg	-0.0508709694839	0.416666666667	36	0	0	-1.762309
wt	12	31	tgctcagat	0.0623540140799	0.346153846154	26	0	0	-1.1077664
wt	12	17	tggtgagca	-0.090319624638	0.416666666667	12	0	0	-0.25859526
mut	12	82	cagtcacag	-0.291165564125	0.441558441558	27	11	15	-1.1389338
mut	12	72	acatgacgg	0.390478992917	0.44776119403	17	11	15	-0.23705118
mut	12	60	ttgtgaggc	-1.03767995916	0.472727272727	5	11	15	-0.028599734
mut	12	41	ccatgagcg	-0.0508709694839	0.416666666667	36	0	0	-1.762309
mut	12	31	tgctcagat	0.0623540140799	0.346153846154	26	0	0	-1.1077664
mut	12	17	tggtgagca	-0.090319624638	0.416666666667	12	0	0	-0.25859526

Variant Effect Predictor tool

ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000269305.9	protein_coding	5/11	630	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000359597.8	protein_coding	4/9	488	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000413465.6	protein_coding	4/7	488	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000420246.6	protein_coding	5/12	621	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000445888.6	protein_coding	5/11	624	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000455263.6	protein_coding	5/12	621	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	downstream_gene_variant	TP53	ENSG00000141510	Transcript	ENST00000503591.1	protein_coding	-	-	-	-	-	-	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000504290.5	protein_coding	1/8	370	92	31	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000504937.5	protein_coding	1/7	370	92	31	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	non_coding_transcript_exon_variant	TP53	ENSG00000141510	Transcript	ENST00000505014.5	retained_intron	4/5	744	-	-	-	-	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000508793.5	protein_coding	5/5	626	488	163	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000509690.5	protein_coding	2/6	224	92	31	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000510385.5	protein_coding	1/8	370	92	31	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	missense_variant	TP53	ENSG00000141510	Transcript	ENST00000514944.5	protein_coding	4/6	288	209	70	Y/C	TAC/TGC	rs148924904, CM942135, COSV52663142, COSV52676381
ENST00000510385.5:c.92A>G	17:7675124-7675124	C	upstream_gene_variant	TP53	ENSG00000141510	Transcript	ENST00000574684.1	processed_transcript	-	-	-	-	-	-	rs148924904, CM942135, COSV52663142, COSV52676381

ESEfinder

Se encuentran solo dos resultados con puntuaciones positivas para las matrices 3’ (483 y 489):

483 (-160)	GCCATCTACAAGCAGTCACAGCACATGACG	-17.54070	483 (-160)	GCCATCTACAAGCAGTCACAGCACATGACG	1.87150	483 (-160)	GCCATCTACAAGCAGTCACAGCACATGACG	-18.28930	483 (-160)	GCCATCTACAAGCAGTCACAGCACATGACG	1.72340
484 (-159)	CCATCTACAAGCAGTCACAGCACATGACGG	-25.51950	484 (-159)	CCATCTACAAGCAGTCACAGCACATGACGG	-30.84620	484 (-159)	CCATCTACAAGCAGTCACAGCACATGACGG	-21.63710	484 (-159)	CCATCTACAAGCAGTCACAGCACATGACGG	-30.24320
485 (-158)	CATCTACAAGCAGTCACAGCACATGACGGA	-20.26110	485 (-158)	CATCTACAAGCAGTCACAGCACATGACGGA	-27.68730	485 (-158)	CATCTACAAGCAGTCACAGCACATGACGGA	-20.84180	485 (-158)	CATCTACAAGCAGTCACAGCACATGACGGA	-29.53490
486 (-157)	ATCTACAAGCAGTCACAGCACATGACGGAG	-29.35710	486 (-157)	ATCTACAAGCAGTCACAGCACATGACGGAG	-31.13760	486 (-157)	ATCTACAAGCAGTCACAGCACATGACGGAG	-25.36780	486 (-157)	ATCTACAAGCAGTCACAGCACATGACGGAG	-33.07360
487 (-156)	TCTACAAGCAGTCACAGCACATGACGGAGG	-29.42000	487 (-156)	TCTACAAGCAGTCACAGCACATGACGGAGG	-10.41070	487 (-156)	TCTACAAGCAGTCACAGCACATGACGGAGG	-27.50370	487 (-156)	TCTACAAGCAGTCACAGCACATGACGGAGG	-10.76860
488 (-155)	CTACAAGCAGTCACAGCACATGACGGAGGT	-4.72970	488 (-155)	CTACAAGCAGTCACAGCACATGACGGAGGT	-31.79980	488 (-155)	CTACAAGCAGTCACAGCACATGACGGAGGT	-4.90460	488 (-155)	CTACAAGCAGTCACAGCACATGACGGAGGT	-33.90360
489 (-154)	TACAAGCAGTCACAGCACATGACGGAGGTT	-21.93590	489 (-154)	TACAAGCAGTCACAGCACATGACGGAGGTT	0.58030	489 (-154)	TACAAGCAGTCACAGCACATGACGGAGGTT	-18.19140	489 (-154)	TACAAGCAGTCACAGCACATGACGGAGGTT	0.14960

Si comparamos con la secuencia mutante, se observa que las puntuaciones has aumentado pero muy levemente:

483 (-160)	GCCATCTGCAAGCAGTCACAGCACATGACG	-17.58150	483 (-160)	GCCATCTGCAAGCAGTCACAGCACATGACG	2.16300	483 (-160)	GCCATCTGCAAGCAGTCACAGCACATGACG	-18.36670	483 (-160)	GCCATCTGCAAGCAGTCACAGCACATGACG	2.07280
484 (-159)	CCATCTGCAAGCAGTCACAGCACATGACGG	-25.36300	484 (-159)	CCATCTGCAAGCAGTCACAGCACATGACGG	-30.35240	484 (-159)	CCATCTGCAAGCAGTCACAGCACATGACGG	-21.47780	484 (-159)	CCATCTGCAAGCAGTCACAGCACATGACGG	-29.55980
485 (-158)	CATCTGCAAGCAGTCACAGCACATGACGGA	-20.09810	485 (-158)	CATCTGCAAGCAGTCACAGCACATGACGGA	-27.06120	485 (-158)	CATCTGCAAGCAGTCACAGCACATGACGGA	-20.63680	485 (-158)	CATCTGCAAGCAGTCACAGCACATGACGGA	-28.73540
486 (-157)	ATCTGCAAGCAGTCACAGCACATGACGGAG	-29.41370	486 (-157)	ATCTGCAAGCAGTCACAGCACATGACGGAG	-30.51760	486 (-157)	ATCTGCAAGCAGTCACAGCACATGACGGAG	-25.43460	486 (-157)	ATCTGCAAGCAGTCACAGCACATGACGGAG	-32.33460
487 (-156)	TCTGCAAGCAGTCACAGCACATGACGGAGG	-29.18660	487 (-156)	TCTGCAAGCAGTCACAGCACATGACGGAGG	-9.94690	487 (-156)	TCTGCAAGCAGTCACAGCACATGACGGAGG	-27.31580	487 (-156)	TCTGCAAGCAGTCACAGCACATGACGGAGG	-10.16100
488 (-155)	CTGCAAGCAGTCACAGCACATGACGGAGGT	-4.48450	488 (-155)	CTGCAAGCAGTCACAGCACATGACGGAGGT	-31.32790	488 (-155)	CTGCAAGCAGTCACAGCACATGACGGAGGT	-4.66980	488 (-155)	CTGCAAGCAGTCACAGCACATGACGGAGGT	-33.27430
489 (-154)	TGCAAGCAGTCACAGCACATGACGGAGGTT	-21.93530	489 (-154)	TGCAAGCAGTCACAGCACATGACGGAGGTT	0.97920	489 (-154)	TGCAAGCAGTCACAGCACATGACGGAGGTT	-18.17790	489 (-154)	TGCAAGCAGTCACAGCACATGACGGAGGTT	0.74380

Por lo tanto, se podría estar produciendo un sitio *acceptor* ligeramente más fuerte pero es poco probable.

En cuanto a los sitios ESE, si comparamos para ambas secuencias (WT arriba, mutante abajo), se ven valores alterados en considerablemente en algunos valores, por lo que podría haber una alteración en estos sitios.:

484 (-159)	CCATCTA	-2.01728	484 (-159)	CCATCTA	-0.24801	484 (-159)	CCATCTAC	-3.86675	484 (-159)	CCATCTA	-2.55818
485 (-158)	CATCTAC	-4.02258	485 (-158)	CATCTAC	-2.08960	485 (-158)	CATCTACA	0.36299	485 (-158)	CATCTAC	0.99753
486 (-157)	ATCTACA	-4.59096	486 (-157)	ATCTACA	-3.38827	486 (-157)	ATCTACAA	1.87565	486 (-157)	ATCTACA	-2.50481
487 (-156)	TCTACAA	-2.97461	487 (-156)	TCTACAA	-2.23269	487 (-156)	TCTACAAG	-3.20104	487 (-156)	TCTACAA	-0.97844
488 (-155)	CTACAAG	-3.66525	488 (-155)	CTACAAG	-1.33051	488 (-155)	CTACAAGC	-3.56460	488 (-155)	CTACAAG	5.08782
489 (-154)	TACAAGC	-1.09177	489 (-154)	TACAAGC	-1.04285	489 (-154)	TACAAGCA	-1.13904	489 (-154)	TACAAGC	-0.74952
490 (-153)	ACAAGCA	-3.13342	490 (-153)	ACAAGCA	-2.03513	490 (-153)	ACAAGCAG	-2.96611	490 (-153)	ACAAGCA	-1.53616

484 (-159)	CCATCTG	-2.74406	484 (-159)	CCATCTG	-0.68338	484 (-159)	CCATCTGC	-5.16365	484 (-159)	CCATCTG	-0.17645
485 (-158)	CATCTGC	-1.44484	485 (-158)	CATCTGC	-0.30994	485 (-158)	CATCTGCA	0.72447	485 (-158)	CATCTGC	1.56536
486 (-157)	ATCTGCA	-2.67134	486 (-157)	ATCTGCA	-1.68188	486 (-157)	ATCTGCAA	1.45038	486 (-157)	ATCTGCA	-5.05937
487 (-156)	TCTGCAA	-5.88318	487 (-156)	TCTGCAA	-3.95245	487 (-156)	TCTGCAAG	-3.20104	487 (-156)	TCTGCAA	-1.13374
488 (-155)	CTGCAAG	-1.60146	488 (-155)	CTGCAAG	0.05324	488 (-155)	CTGCAAGC	-4.86149	488 (-155)	CTGCAAG	2.47903
489 (-154)	TGCAAGC	-1.53899	489 (-154)	TGCAAGC	-0.74463	489 (-154)	TGCAAGCA	-0.77819	489 (-154)	TGCAAGC	0.69564
490 (-153)	GCAAGCA	-2.20554	490 (-153)	GCAAGCA	-1.79758	490 (-153)	GCAAGCAG	-1.21404	490 (-153)	GCAAGCA	-2.99088

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	18	13	6	94	1577.7853	45	-61.2266	24	23	144	2059.5707	138	195.4978	176	329	0.53
mut	18	13	6	94	1577.7853	45	-61.2266	25	22	140	1982.8706	141	197.6182	176	328	0.54

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

HOT-SKIP

gaccagcctgggtaacatgatgaaacctcgtctctacaaaaaaatacaaaaaattagccaggcatggtggtgcacaccta
tagtcccagccacttaggaggctgaggtgggaagatcactTGAGGCCAGGAGATGGAGGCTGCAGTGAGCTGTGATCACA
CCACTGTGCTCCAGCCTGAGTGACAGAGCAAGACCCTATCTCAAAAAAAAAAAAAAAAAAAGAAAAGCTCCTGAGGTGTAG
ACGCCAACTCTCTCTAGCTCGCTAGTGGGTTGCAGGAGGTGCTTACGCATGTTTGTTTCTTTGCTGCCGTCTTCCAGTTG
CTTTATCTGTTCACTTGTGCCCTGACTTTCAACTCTGTCTCCTTCCTCTTCCTACAGTACTCCCCTGCCCTCAACAAGAT
GTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCA
TGGCCATCTACAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGAT
Ggtgagcagctggggctggagagacgacagggctggttgcccagggtccccaggcctctgattcctcactgattgctctt
ag