Ejemplo comparación de resultados predictores in sillico

Cambio de estudio MAP4K4 c.640-1G>T (chr2:101834408 G/T o NM_001242559.2:c.640-1G>T)

Exón 8 e intrones adyacentes:

```
attcatagatgtgagtggttttaattaaatcaattttctccttccttcctttccttcctt
tccttccctcccttccctccctcccctccctccatccctccatccctccatccctccatc
cctcccttcctcccctcccctcccctccgctccgtaagttctacaaatgtgcttgtacttt
ttaattttctggcaaatttccttttggtttatatagcaaagttgtgtgaatacccagaca
tg    a    g    t    a c ac c a  at t   cgt  t    a    gc g
AGTGATCTTTGGTCTTGTGGCATTACAGCCATTGAGATGGCAGAAGGTGCTCCCC
gtaagtaactttcttttcttttttagctcacttgttacatgtgacttaaaccctcccaag
acttcaaaagaacagctcagctccatggataaaggcatttctggtggactgtttagatg
aaggcaacataaatatagtcaagtttttagatgtagaaggaattcttggcaaccactatta
ctattccttttgaatatctctaggaggcctataggctacattttctttattaagttgt
actcacatgtgtgacctgaagtagtgctgagatcagataataagtagtaagttttcttttt
```

El cambio se encuentra en la posición justo antes del exón 8 (la **g** 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                          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'
            11         -      +        0.00     ATTCATAGAT^GTGAGTGGTT              11         -      +        0.00     ATTCATAGAT^GTGAGTGGTT
           154         0      +        0.47     CCTCCGCTCC^GTAAGTTCTA             154         0      +        0.47     CCTCCGCTCC^GTAAGTTCTA
           242         2      +        0.47     ACCCAGACAT^GTAAGTTAGT             242         2      +        0.47     ACCCAGACAT^GTAAGTTAGT
           356         1      +        0.63     GGTGCTCCCC^GTAAGTAACT             356         1      +        0.63     GGTGCTCCCC^GTAAGTAACT
           642         -      +        0.00     ATAATAAGTA^GTAAGTTTTC             642         -      +        0.00     ATAATAAGTA^GTAAGTTTTC

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'
      601         55         0      -        0.00     GGTCACACAT^GTGAGTACAA           601         55         0      -        0.00     GGTCACACAT^GTGAGTACAA
      385        271         1      -        0.36     CATGTAACAA^GTGAGCTAAA           385        271         1      -        0.36     CATGTAACAA^GTGAGCTAAA

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'
           300         0      +        0.43     ATTTTTGCAG^AGTGATCTTT           328         1      +        0.16     GGCATTACAG^CCATTGAGAT
           328         1      +        0.16     GGCATTACAG^CCATTGAGAT           380         2      +        0.25     TTCTTTTTAG^CTCACTTGTT
           380         2      +        0.25     TTCTTTTTAG^CTCACTTGTT           591         2      +        0.00     TTTATTTAAG^TTGTACTCAC
           591         2      +        0.00     TTTATTTAAG^TTGTACTCAC
```

Se pierde un sitio *acceptor* en la hebra directa (en rojo) en la secuencia mutante. Coincide con el sitio *acceptor* del exón. Por lo tanto, es probable que esté afectando al *splicing*. Dados los niveles de *confidence* de los otros *acceptor* y sus posiciones, es probable que se use el *acceptor* del siguiente exón, produciéndose *exon skipping.*

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.15.54.30613.0 :

Start	End	Score	Exon	Intron
4	18	0.90	catagat**gt**gagtgg	
147	161	0.98	ccgctcc**gt**aagttc	
235	249	0.97	cagacat**gt**aagtta	
349	363	0.92	gctcccc**gt**aagtaa	
635	649	0.98	ataagta**gt**aagttt	

Donor site predictions for 85.53.15.54.30630.0 :

Start	End	Score	Exon	Intron
4	18	0.90	catagat**gt**gagtgg	
147	161	0.98	ccgctcc**gt**aagttc	
235	249	0.97	cagacat**gt**aagtta	
349	363	0.92	gctcccc**gt**aagtaa	
635	649	0.98	ataagta**gt**aagttt	

Acceptor site predictions for 85.53.15.54.30613.0 :

Start	End	Score	Intron	Exon
196	236	0.76	atttccttttggtttatat**ag**caaagttgtgtgaataccca	
280	320	0.97	cgtactgttttattttgc**ag**agtgatctttggtcttgtgg	
308	348	0.46	tttggtcttgtggcattac**ag**ccattgagatggcagaaggt	
360	400	1.00	gtaactttcttttctttt**ag**ctcacttgttacatgtgact	
538	578	0.78	attccttttgaatatctct**ag**gaggcctataggctacattt	
571	611	0.94	ctacattttctttattta**ag**ttgtactcacatgtgtgacc	

Acceptor site predictions for 85.53.15.54.30630.0 :

Start	End	Score	Intron	Exon
196	236	0.76	atttccttttggtttatat**ag**caaagttgtgtgaataccca	
282	322	0.74	tactgttttattttgcat**ag**tgatctttggtcttgtggca	
308	348	0.46	tttggtcttgtggcattac**ag**ccattgagatggcagaaggt	
360	400	1.00	gtaactttcttttctttt**ag**ctcacttgttacatgtgact	
538	578	0.78	attccttttgaatatctct**ag**gaggcctataggctacattt	
571	611	0.94	ctacattttctttattta**ag**ttgtactcacatgtgtgacc	

El cambio que hay entre ambos es que uno de los *acceptor* (en rojo) en la secuencia WT no aparece en la otra predicción y se activa otro en la mutante dos nucleótidos más a la derecha. Si se empleara este nuevo, se perderían dos nucleótidos del exón. Si este no es viable, se perdería el exón completo.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ttgca(g/t)agtga	gcagag	gcatag	35903	72%

Human Splicing Finder

Broken WT Acceptor Site	Alteration of the WT Acceptor site, most probably affecting splicing			
Algorithm/Matix	**position**	**sequences**	**variation**	
MaxEnt Acceptor site	chr2:101834389	- REF : GTACTGTTTTATTTTTGCAGAGT - ALT : GTACTGTTTTATTTTTGCATAGT	9.2 > 0.6 => -93.48%	
HSF Acceptor site (matrix AG)	chr2:101834397	- REF : TTATTTTTGCAGAG - ALT : TTATTTTTGCATAG	92.36 > 64.49 => -30.18%	

New Acceptor splice site	Activation of a cryptic Acceptor site. Potential alteration of splicing			
Algorithm/Matix	**position**	**sequences**	**variation**	
MaxEnt Acceptor site	chr2:101834391	- REF : ACTGTTTTATTTTTGCAGAGTGA - ALT : ACTGTTTTATTTTTGCATAGTGA	-13.08 > 6.43 => 149.16%	

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	31	388	ccgtaagtt	-0.276080102925	0.537859007833	13	21	35	-0.029431445
wt	31	362	cttttaatt	-4.22449622523	0.532212885154	2	6	17	-1.0486502
wt	31	361	ttttaattt	-0.995102413301	0.533707865169	1	6	17	0.27959129
wt	31	334	ggtttatat	-3.56556400653	0.519756838906	43	15	25	-3.3153744
wt	31	316	gtgtgaata	-1.25998720305	0.527331189711	25	15	25	-1.2708161
wt	31	300	atgtaagtt	-1.19087085871	0.532203389831	9	15	25	-0.22940704
wt	31	296	aagttagtg	-3.29097962164	0.53264604811	5	15	25	-0.79836261
wt	31	266	ctgtaacgt	2.32997371862	0.524904214559	5	15	28	1.4279534
wt	31	254	gttttattt	-4.38371630518	0.518072289157	1	7	15	-1.0708926
wt	31	240	==gagtgatct==	0.32003046975	0.51914893617	1	13	22	==0.83640509==
wt	31	220	gcattacag	-1.23893443579	0.506976744186	23	10	15	-1.2357086
wt	31	210	cattgagat	-0.784144460909	0.507317073171	13	10	15	-0.4245437
wt	31	186	ccgtaagta	-0.580073238124	0.519337016575	5	27	55	0.53826227
wt	31	182	aagtaactt	0.958607225608	0.525423728814	1	27	55	1.3958875
wt	31	166	tttttagct	-3.58154174642	0.484472049689	2	10	17	-0.81232202
wt	31	161	agctcactt	0.969306583605	0.480769230769	143	14	18	-7.9473931
wt	31	154	ttgttacat	-1.61570658262	0.469798657718	136	14	18	-8.5200029
wt	31	147	atgtgactt	0.642799644865	0.471830985915	129	14	18	-7.1919463
wt	31	143	gacttaaac	-1.24917643313	0.463768115942	125	14	18	-7.6821552
wt	31	142	acttaaacc	-1.1071468731	0.467153284672	124	14	18	-7.5621523
wt	31	124	acttcaaaa	-3.17258789262	0.436974789916	106	14	18	-7.241247
wt	31	110	agctcagct	0.337745592429	0.457142857143	92	14	18	-4.9740954
wt	31	97	ggataaagg	-1.83816607884	0.45652173913	79	14	18	-5.0033901
wt	31	74	tgtttagat	-3.8051613457	0.434782608696	56	14	18	-4.3247213
wt	31	69	agatgaagg	-2.2566557852	0.453125	51	14	18	-3.3959926
wt	31	58	acataaata	-2.53331538046	0.490566037736	40	14	18	-2.795944
wt	31	49	tagtcaagt	-2.21817162867	0.5	31	14	18	-2.0998186
wt	31	42	gttttagat	-4.75296554591	0.486486486486	24	14	18	-2.6535864

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
mut	31	388	ccgtaagtt	-0.276080102925	0.54046997389	13	21	35	-0.028588193
mut	31	362	cttttaatt	-4.22449622523	0.535014005602	2	6	17	-1.0477455
mut	31	361	ttttaattt	-0.995102413301	0.536516853933	1	6	17	0.28049851
mut	31	334	ggtttatat	-3.56556400653	0.522796352584	43	15	25	-3.3143927
mut	31	316	gtgtgaata	-1.25998720305	0.530546623794	25	15	25	-1.2697776
mut	31	300	atgtaagtt	-1.19087085871	0.535593220339	9	15	25	-0.22831222
mut	31	296	aagttagtg	-3.29097962164	0.536082474227	5	15	25	-0.79725276
mut	31	266	ctgtaacgt	2.32997371862	0.528735632184	5	15	28	1.4291909
mut	31	254	gttttattt	-4.38371630518	0.522088353414	1	7	15	-1.0695956
mut	31	240	==tagtgatct==	0.310477296786	0.51914893617	1	13	22	==0.83266456==
mut	31	220	gcattacag	-1.23893443579	0.506976744186	23	10	15	-1.2357086
mut	31	210	cattgagat	-0.784144460909	0.507317073171	13	10	15	-0.4245437
mut	31	186	ccgtaagta	-0.580073238124	0.519337016575	5	27	55	0.53826227
mut	31	182	aagtaactt	0.958607225608	0.525423728814	1	27	55	1.3958875
mut	31	166	tttttagct	-3.58154174642	0.484472049689	2	10	17	-0.81232202
mut	31	161	agctcactt	0.969306583605	0.480769230769	143	14	18	-7.9473931
mut	31	154	ttgttacat	-1.61570658262	0.469798657718	136	14	18	-8.5200029
mut	31	147	atgtgactt	0.642799644865	0.471830985915	129	14	18	-7.1919463
mut	31	143	gacttaaac	-1.24917643313	0.463768115942	125	14	18	-7.6821552
mut	31	142	acttaaacc	-1.1071468731	0.467153284672	124	14	18	-7.5621523
mut	31	124	acttcaaaa	-3.17258789262	0.436974789916	106	14	18	-7.241247
mut	31	110	agctcagct	0.337745592429	0.457142857143	92	14	18	-4.9740954
mut	31	97	ggataaagg	-1.83816607884	0.45652173913	79	14	18	-5.0033901
mut	31	74	tgtttagat	-3.8051613457	0.434782608696	56	14	18	-4.3247213
mut	31	69	agatgaagg	-2.2566557852	0.453125	51	14	18	-3.3959926
mut	31	58	acataaata	-2.53331538046	0.490566037736	40	14	18	-2.795944
mut	31	49	tagtcaagt	-2.21817162867	0.5	31	14	18	-2.0998186
mut	31	42	gttttagat	-4.75296554591	0.486486486486	24	14	18	-2.6535864

El único cambio que hay es en el resultado donde se encuentra la mutación. La puntuación sigue siendo la misma, por lo que sigue siendo un BP de interés, no afectando al *splicing.*

Variant Effect Predictor tool

ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	intron_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000302217.9	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000324219.8	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000347699.8	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000350198.8	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000350878.8	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000413150.6	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000417294.5	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000425019.5	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	intron_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000456652.5	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant, non_coding_transcript_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000476609.1	retained_intron	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	downstream_gene_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000489490.1	retained_intron	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	downstream_gene_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000496989.5	processed_transcript	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000625522.2	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	intron_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000627726.2	protein_coding	-
ENST00000324219.8:c.640-1G>T	2:101834408-101834408	T	splice_acceptor_variant	MAP4K4	ENSG00000071054	Transcript	ENST00000634702.1	protein_coding	-

Se trata de una mutación que afecta al *splicing*, concretamente al *acceptor* del exón.

ESEfinder

Solo se obtiene un resultado positivo con la posición de interés en las matrices 3'SS:

| 286 (-370) | gttttatttttgcagAGTGATCTTTGGTCT | -27.37730 | | 286 (-370) | gttttatttttgcagAGTGATCTTTGGTCT | 8.50350 | | 286 (-370) | gttttatttttgcagAGTGATCTTTGGTCT | -25.81740 | | 286 (-370) | gttttatttttgcagAGTGATCTTTGGTCT | 8.31270 |

Sin embargo, estas puntuaciones en la secuencia mutante son negativas:

| 286 (-370) | gttttatttttgcatAGTGATCTTTGGTCT | -31.34710 | | 286 (-370) | gttttatttttgcatAGTGATCTTTGGTCT | -3.58410 | | 286 (-370) | gttttatttttgcatAGTGATCTTTGGTCT | -29.74060 | | 286 (-370) | gttttatttttgcatAGTGATCTTTGGTCT | -4.64070 |

Por lo tanto, la mutación está generando la pérdida de un sitio *acceptor,* afectando al *splicing.*