Ejemplo comparación de resultados predictores in sillico

Cambio de estudio PIK3CD c.2808C>T (chr1:9724365 C/T, rs11121484 o NM_005026.5:c.2808C>T)

Exón 22 e intrones adyacentes:

```
gtacaggggctggtgctggcggctgctgtggggacttggcttctggccccagcctgctgg
cccctctgcctagcacacagctctgtggcaggggtccccagccctgctggcttcctgtc
tccctggattctctcctgtctgacaccttctcaatcctcccccctcctctcccctcccct
cag
CTGTTCCACATTGATTTTGGCCACTTTCTGGGGAATTTCAAGACCAAGTTTGGAATCAAC
CGCCGAGCGTGTCCCATTCATCCTCACCCTACGACTTTGTCCATGTGATTCAGCAGGGGAAG
ACTAATAATAGTGAGAAATTTGAACG
gtgagagtgcctgagcccaccagatgcccctcggtgtgggggcccagggaacaggggcag
aggttccaggcaggtgcaggatgggctcaggtctcaacccacacctggccctcac
cccaactgttgatgggtttggaacatgccctgctccaccctgcagtgcccttttgggc
aatgtgggcaggtttgtgggtcatgtagcaggaggctggctggggcacggggtcagtta
gcagaactggaggccttgtgtccacccattatcagggcaagggcaggtgtccttgggaa
ggggctggttggatgcagagcggccctctggcctgtggctgggagttcccagagcctcac
ttcctctgtccctacctgcag
```

El cambio se encuentra en segunda línea del exón 22 (la **c** en color verde pegada a la **g** amarilla).

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'
            314        1      +        0.31      ACTAATAATA^GTGAGAAATT
            330        2      +        0.95      AATTTGAACG^GTGAGAGTGC H

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'
            183        0      +        1.00      CTCCCCTCAG^CTGTTCCACA H
            225        0      +        0.18      GAATTTCAAG^ACCAAGTTTG
            231        0      +        0.07      CAAGACCAAG^TTTGGAATCA
            495        1      +        0.23      CACCCTGCAG^TGCCCCTTTT

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon      3'
      305        407         2      -        0.43      CTATTATTAG^TCTTCCCCTG
      276        436         1      -        0.17      ATGGACAAAG^TCGTAGGTGA
```

```
Donor splice sites, direct strand
---------------------------------
         pos 5'->3'  phase strand  confidence  5'      exon intron      3'
            314        1      +        0.24      ACTAATAATA^GTGAGAAATT
            330        2      +        0.95      AATTTGAACG^GTGAGAGTGC H

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'
            183        0      +        0.97      CTCCCCTCAG^CTGTTCCACA H
            225        0      +        0.17      GAATTTCAAG^ACCAAGTTTG
            495        1      +        0.23      CACCCTGCAG^TGCCCCTTTT

Acceptor splice sites, complement strand
----------------------------------------
   pos 3'->5'  pos 5'->3'  phase strand  confidence  5'    intron exon      3'
      305        407         2      -        0.34      CTATTATTAG^TCTTCCCCTG
      276        436         1      -        0.18      ATGGACAAAG^TCATAGGTGA
      270        442         1      -        0.07      AAAGTCATAG^GTGAGGATGA
```

Aparición nuevo sitio de *splicing* en la secuencia mutante. Tiene muy poca *confidence* y está en la hebra reversa, por lo que es probable que no afecte al *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.15.54.21973.0 :

Start	End	Score	Exon	Intron
323	337	0.96	ttgaacg**gt**gagagt	
514	528	0.50	tgggcag**gt**ttgtgg	
555	569	0.40	cacgggg**gt**cagtta	

Donor site predictions for 85.53.15.54.22001.0 :

Start	End	Score	Exon	Intron
323	337	0.96	ttgaacg**gt**gagagt	
514	528	0.50	tgggcag**gt**ttgtgg	
555	569	0.40	cacgggg**gt**cagtta	

Acceptor site predictions for 85.53.15.54.21973.0 :

Start	End	Score	Intron	Exon
53	93	0.53	cctgctggcccctctgcct**ag**cacacagctctgtggcaggg	
163	203	0.94	cctcctctcccctcccctc**ag**ctgttccacattgattttgg	
277	317	0.67	tttgtccatgtgattcagc**ag**gggaagactaataatagtga	
475	515	0.68	tgccctgctccaccctgc**ag**tgcccctttttgggcaatgtg	

Acceptor site predictions for 85.53.15.54.22001.0 :

Start	End	Score	Intron	Exon
53	93	0.53	cctgctggcccctctgcct**ag**cacacagctctgtggcaggg	
163	203	0.94	cctcctctcccctcccctc**ag**ctgttccacattgattttgg	
277	317	0.67	tttgtccatgtgattcagc**ag**gggaagactaataatagtga	
475	515	0.68	tgccctgctccaccctgc**ag**tgcccctttttgggcaatgtg	

No hay cambios.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
accta(c/t)gactt	cgactt	tgactt	30563	78%
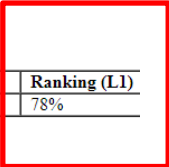
CRYP-SKIP

Exon length (bp)	146
PESS (<=-2.62) density	0.00
NN 5'ss score density	0.21
SF2/ASF score density	8.39
FAS-ESS (hex2) density	4.11
EIE score density	398.66
Probability of cryptic splice site activation (P_{CR-E})	0.61
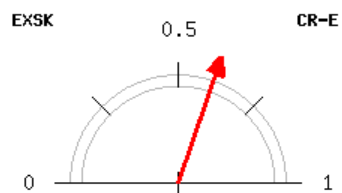
Parece que hay un sitio críptico de *splicing* dentro del propio exón, pero el cambio de no lo toma en consideración, por lo que no debe considerar que tenga algún efecto en el *splicing*.

>wt

```
                                                              0.02
ctcctctcccctcccctcagCTGTTCCACATTGATTTTGGCCACTTTCTGGGGAATTTCAAGACCAAGTTTGGAATCAAC
                                                    0.49
                                0.10
CGCGAGCGTGTCCCATTCATCCTCACCTATGACTTTGTCCATGTGATTCAGCAGGGGAAGACTAATAATAGTGAGAAATT
TGAACGgtgagag
```
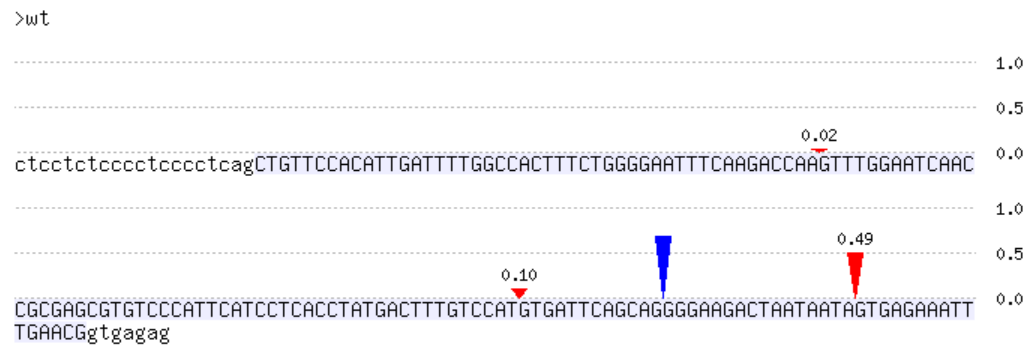
Human Splicing Finder

⊖ Alteration of auxiliary sequences	Significant alteration of ESE / ESS motifs ratio (-4)	
Algorithm/Matix	**position**	**sequence**
ESE_ASF (ESE Site Broken)	chr1:9724361	CCTACGA
ESE_ASFB (ESE Site Broken)	chr1:9724361	CCTACGA
Sironi_motif3 (ESS Site Broken)	chr1:9724361	CCTACGAC
ESE_SRp55 (ESE Site Broken)	chr1:9724363	TACGAC
RESCUE ESE (ESE Site Broken)	chr1:9724364	ACGACT
IIE (New ESS Site)	chr1:9724365	TGACTT

SVM-BPfinder

```
seq_id  agez  ss_dist  bp_seq     bp_scr            y_cont          ppt_off  ppt_len  ppt_scr  svm_scr
wt      14    444      gtctgacac  2.99502765393     0.521640091116  1        48       81       2.4342323
wt      14    434      ttctcaatc  0.388726718022    0.51515151552   2        37       68       1.2272424
wt      14    406      ccctcagct  1.29086331147     0.486284289277  2        9        13       1.0587768
wt      14    391      cattgattt  0.0540481574368   0.481865284974  7        9        18       0.30316561
wt      14    365      atttcaaga  -3.07656471962    0.472222222222  29       22       31       -2.1971888
wt      14    347      gaatcaacc  -1.4716882453     0.476608187135  11       22       31       -0.42801751
wt      14    326      cattcatcc  0.139031874197    0.470404984424  1        11       17       0.70321349
wt      14    320      tcctcacct  2.41413148315     0.463492063492  8        10       17       1.1487016
wt      14    299      atgtgattc  -0.387148056259   0.452380952381  135      29       34       -7.8322377
wt      14    295      gattcagca  -0.777165203275   0.448275862069  131      29       34       -7.7330804
wt      14    280      gactaataa  0.934035186093    0.461818181818  116      29       34       -6.109216
wt      14    277      taataatag  -0.0249293398427  0.463235294118  113      29       34       -6.2943433
wt      14    271      tagtgagaa  -1.85680511767    0.466165413534  107      29       34       -6.630873
wt      14    262      atttgaacg  -0.636856455934   0.470817120623  98       29       34       -5.5820181
wt      14    255      cggtgagag  -0.236661987024   0.476           91       29       34       -4.9805609
wt      14    245      gcctgagcc  1.37313434892     0.479166666667  81       29       34       -3.7162436
wt      14    167      ggctcaggt  -0.452459714721   0.506172839506  3        29       34       0.51494126
wt      14    160      gtctcaacc  0.749798321109    0.503225806452  2        23       28       0.99213354
wt      14    140      ccctcaccc  3.06645120319     0.474074074074  27       19       26       0.28870782
wt      14    127      tgttgatgg  1.16038468823     0.459016393443  14       19       26       0.36040762
wt      14    56       gggtcatgt  -0.680573548496   0.333333333333  51       0        0        -2.9852563
wt      14    23       gggtcagtt  -2.03352206854    0.333333333333  18       0        0        -1.426155
wt      14    19       cagttagca  -2.60979930003    0.285714285714  14       0        0        -1.4139811
```

```
seq_id  agez  ss_dist  bp_seq     bp_scr            y_cont          ppt_off  ppt_len  ppt_scr  svm_scr
mut     14    444      gtctgacac  2.99502765393     0.521640091116  1        48       81       2.4342323
mut     14    434      ttctcaatc  0.388726718022    0.51515151552   2        37       68       1.2272424
mut     14    406      ccctcagct  1.29086331147     0.486284289277  2        9        13       1.0587768
mut     14    391      cattgattt  0.0540481574368   0.481865284974  7        9        18       0.30316561
mut     14    365      atttcaaga  -3.07656471962    0.472222222222  29       22       32       -2.187873
mut     14    347      gaatcaacc  -1.4716882453     0.476608187135  11       22       32       -0.41870167
mut     14    326      cattcatcc  0.139031874197    0.470404984424  1        11       18       0.71252933
mut     14    320      tcctcacct  2.41413148315     0.463492063492  8        10       17       1.1487016
mut     14    313      ctatgactt  0.484197118369    0.461038961039  1        10       17       0.83533704
mut     14    299      atgtgattc  -0.387148056259   0.452380952381  135      29       34       -7.8322377
mut     14    295      gattcagca  -0.777165203275   0.448275862069  131      29       34       -7.7330804
mut     14    280      gactaataa  0.934035186093    0.461818181818  116      29       34       -6.109216
mut     14    277      taataatag  -0.0249293398427  0.463235294118  113      29       34       -6.2943433
mut     14    271      tagtgagaa  -1.85680511767    0.466165413534  107      29       34       -6.630873
mut     14    262      atttgaacg  -0.636856455934   0.470817120623  98       29       34       -5.5820181
mut     14    255      cggtgagag  -0.236661987024   0.476           91       29       34       -4.9805609
mut     14    245      gcctgagcc  1.37313434892     0.479166666667  81       29       34       -3.7162436
mut     14    167      ggctcaggt  -0.452459714721   0.506172839506  3        29       34       0.51494126
mut     14    160      gtctcaacc  0.749798321109    0.503225806452  2        23       28       0.99213354
mut     14    140      ccctcaccc  3.06645120319     0.474074074074  27       19       26       0.28870782
mut     14    127      tgttgatgg  1.16038468823     0.459016393443  14       19       26       0.36040762
mut     14    56       gggtcatgt  -0.680573548496   0.333333333333  51       0        0        -2.9852563
mut     14    23       gggtcagtt  -2.03352206854    0.333333333333  18       0        0        -1.426155
mut     14    19       cagttagca  -2.60979930003    0.285714285714  14       0        0        -1.4139811
```

Aparece un BP nuevo con puntuación positiva en la secuencia mutante que coincide con la mutación de estudio. Por lo tanto, podría considerarlo como el final del intrón y se podría perder la parte inicial del exón (95 pb).

Variant Effect Predictor tool

Variant	Location	Allele	Consequence	Symbol	Gene	Feature	Biotype	EXON	cDNA pos	CDS pos	Protein pos	Y	Codons	Existing var
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	synonymous_variant	PIK3CD	ENSG00000171608 Transcript	ENST00000361110.6	protein_coding	21/23	2995	2880	960	Y	TAC/TAT	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	downstream_gene_variant	CLSTN1	ENSG00000171603 Transcript	ENST00000361311.4	protein_coding	-	-	-	-	-	-	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	downstream_gene_variant	CLSTN1	ENSG00000171603 Transcript	ENST00000377298.9	protein_coding	-	-	-	-	-	-	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	synonymous_variant	PIK3CD	ENSG00000171608 Transcript	ENST00000377346.9	protein_coding	22/24	3017	2808	936	Y	TAC/TAT	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	downstream_gene_variant	CLSTN1	ENSG00000171603 Transcript	ENST00000435891.5	protein_coding	-	-	-	-	-	-	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	downstream_gene_variant	CLSTN1	ENSG00000171603 Transcript	ENST00000477264.1	processed_transcript	-	-	-	-	-	-	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	synonymous_variant	PIK3CD	ENSG00000171608 Transcript	ENST00000536656.5	protein_coding	23/25	3088	2880	960	Y	TAC/TAT	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	synonymous_variant	PIK3CD	ENSG00000171608 Transcript	ENST00000543390.2	protein_coding	22/24	2995	2880	960	Y	TAC/TAT	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	synonymous_variant	PIK3CD	ENSG00000171608 Transcript	ENST00000628140.2	protein_coding	22/24	3088	2880	960	Y	TAC/TAT	rs11121484
ENST00000377346.8:c.2808C>T	1:9724365-9724365	T	downstream_gene_variant	CLSTN1	ENSG00000171603 Transcript	ENST00000650348.1	nonsense_mediated_decay	-	-	-	-	-	-	rs11121484

ESEfinder

Aparece un resultado en la búsqueda de sitios de *splicing* con la posición de interés con puntuaciones positivas en las matrices de 5'SS:

| 271 (-441) | TACGACTTTGTCCATGTGATTCAGCAGGGG | 4.58580 | 271 (-441) | TACGACTTTGTCCATGTGATTCAGCAGGGG | -4.39310 | 271 (-441) | TACGACTTTGTCCATGTGATTCAGCAGGGG | 4.85540 | 271 (-441) | TACGACTTTGTCCATGTGATTCAGCAGGGG | -5.79040 |

Cuando buscamos el resultado equivalente para la secuencia mutante, las puntuaciones se reducen un poco:

| 271 (-441) | TATGACTTTGTCCATGTGATTCAGCAGGGG | 3.84280 | 271 (-441) | TATGACTTTGTCCATGTGATTCAGCAGGGG | -4.02550 | 271 (-441) | TATGACTTTGTCCATGTGATTCAGCAGGGG | 4.07190 | 271 (-441) | TATGACTTTGTCCATGTGATTCAGCAGGGG | -5.49100 |

Puede que se esté debilitando el sitio 5'SS y esté afectando al *splicing*.

Cuando realizamos la búsqueda de ESE, para los resultados donde se encuentra la posición de interés, solo se obtienen dos resultados con puntuaciones positivas en más de una matriz (269 y 273):

267 (-445)	CACCTAC	-1.71256	267 (-445)	CACCTAC	-0.02073	267 (-445)	CACCTACG	1.38182	267 (-445)	CACCTAC	-0.35075
268 (-444)	ACCTACG	-5.92067	268 (-444)	ACCTACG	-4.07242	268 (-444)	ACCTACGA	-0.80965	268 (-444)	ACCTACG	0.18898
269 (-443)	CCTACGA	2.55362	269 (-443)	CCTACGA	2.65201	269 (-443)	CCTACGAC	-5.07860	269 (-443)	CCTACGA	-0.76730
270 (-442)	CTACGAC	-3.22549	270 (-442)	CTACGAC	-0.61315	270 (-442)	CTACGACT	-2.31430	270 (-442)	CTACGAC	1.69063
271 (-441)	TACGACT	-4.72468	271 (-441)	TACGACT	-3.21321	271 (-441)	TACGACTT	-1.38956	271 (-441)	TACGACT	-2.74954
272 (-440)	ACGACTT	-0.36105	272 (-440)	ACGACTT	-0.56714	272 (-440)	ACGACTTT	-6.33139	272 (-440)	ACGACTT	-5.04430
273 (-439)	CGACTTT	-2.15869	273 (-439)	CGACTTT	-0.05448	273 (-439)	CGACTTTG	0.92421	273 (-439)	CGACTTT	0.49615

Si buscamos las predicciones equivalentes para la secuencia mutante, se observa que para 269 las puntuaciones se reducen considerablemente y para 273 se mantienen parecidas.

Pos	Seq	Value
267 (-445)	CACCTAT	0.14084
268 (-444)	ACCTATG	-5.46296
269 (-443)	CCTATGA	0.02989
270 (-442)	CTATGAC	-4.68281
271 (-441)	TATGACT	-7.03470
272 (-440)	ATGACTT	0.24188
273 (-439)	TGACTTT	-5.10918

Pos	Seq	Value
267 (-445)	CACCTAT	1.17597
268 (-444)	ACCTATG	-3.96842
269 (-443)	CCTATGA	0.84022
270 (-442)	CTATGAC	-1.86790
271 (-441)	TATGACT	-5.28207
272 (-440)	ATGACTT	-0.31836
273 (-439)	TGACTTT	-3.15954

Pos	Seq	Value
267 (-445)	CACCTATG	2.02625
268 (-444)	ACCTATGA	-2.53866
269 (-443)	CCTATGAC	-5.83784
270 (-442)	CTATGACT	-2.54514
271 (-441)	TATGACTT	-1.95849
272 (-440)	ATGACTTT	-4.94299
273 (-439)	TGACTTTG	0.90308

Pos	Seq	Value
267 (-445)	CACCTAT	-1.88985
268 (-444)	ACCTATG	-1.52795
269 (-443)	CCTATGA	0.72395
270 (-442)	CTATGAC	-0.69401
271 (-441)	TATGACT	-1.40127
272 (-440)	ATGACTT	-5.35635
273 (-439)	TGACTTT	0.85285

Por lo tanto, es probable que se esté debilitando el ESE 269, alterando al *splicing.*

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	6	4	39	605.0194	14	-26.7583	4	14	48	598.9967	35	50.7792	63	101	0.62
mut	0	6	4	40	619.0086	14	-26.7583	4	13	48	575.1791	33	48.4335	64	98	0.65

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

HOT-SKIP

>wt
gtacaggggctggtgctggcggctgctgtggggacttggcttctggccccagcctgctggcccctctgcctagcacacag
ctctgtggcaggggtcccccagccctgctggcttcctgtctcccctggattctctcctgtctgacaccttctcaatcctc
cccctcctctcccctcccctcagCTGTTCCACATTGATTTTGGCCACTTTCTGGGGAATTTCAAGACCAAGTTTGGAATC
AACCGCGAGCGTGTCCCATTCATCCTCACCTACGACTTTGTCCATGTGATTCAGCAGGGGAAGACTAATAATAGTGAGAA
ATTTGAACGgtgagagtgcctgagcccccaccagatgcccctcggtgtggggccccagggaacagggcagaggttcccagg
cagggtgcaggatggggctcaggtctcaaccccacacctggcccctcaccccaactgttgatgggtttggaacatgcccc
tgctccaccctgcagtgccccttttgggcaatgtgggcaggtttgtgggtcatgtagcaggaggctggctggggcacggg
ggtcagttagcagaactggaggccttgtgtccacccattatcagggcaagggcaggtgtccttggggaaggggctggttg
gatgcagagcggccctctggcctgtggctgggagttcccagagcctcacttcctctgtcccctacctgcag

>mut
gtacaggggctggtgctggcggctgctgtggggacttggcttctggccccagcctgctggcccctctgcctagcacacag
ctctgtggcaggggtcccccagccctgctggcttcctgtctcccctggattctctcctgtctgacaccttctcaatcctc
cccctcctctcccctcccctcagCTGTTCCACATTGATTTTGGCCACTTTCTGGGGAATTTCAAGACCAAGTTTGGAATC
AACCGCGAGCGTGTCCCATTCATCCTCACCTATGACTTTGTCCATGTGATTCAGCAGGGGAAGACTAATAATAGTGAGAA
ATTTGAACGgtgagagtgcctgagcccccaccagatgcccctcggtgtggggccccagggaacagggcagaggttcccagg
cagggtgcaggatggggctcaggtctcaaccccacacctggcccctcaccccaactgttgatgggtttggaacatgcccc
tgctccaccctgcagtgccccttttgggcaatgtgggcaggtttgtgggtcatgtagcaggaggctggctggggcacggg
ggtcagttagcagaactggaggccttgtgtccacccattatcagggcaagggcaggtgtccttggggaaggggctggttg
gatgcagagcggccctctggcctgtggctgggagttcccagagcctcacttcctctgtcccctacctgcag