

Ejemplo comparación de resultados predictores in silico

Cambio de estudio U2AF1 c.44+1G>A (chr21:43107450 G/A, COSV52347342 o NM_006758.3: c.44+1G>A)

Exón 1 e intrones adyacentes:

```
ccaggcggctctgtggagaagccctcggccccgggttcgggtctgcactgccagcccca  
gtcccggcgcgagggatccagcaggccagcgccgggcgggcgagaaggcgagagagct  
agggccggcggggagggcacggcagggggcgcggggtcccaggaccgaaaaggcggag  
cctccacgcgcactacgtgctccctacgcgcacatgcgtcccgccacgcgcaggcagcac  
ctcgctgactgggttagcgctgcacgtcggggcggggaagagcgtcgtcgcgtccgggt  
GACGTCTCCCGAGGGCGTCGGCAGGGTCGGCGGGTCGGCAGCAGTGTGACGGCAGCGG  
CGGCGGCGGGTGGAAATGGCGGAGTATCTGGCTCCATCTTCGGCACCGAGAAAGACAA  
gtgagtggagccccccgcgggggttgggcgcgatcgggcgcgacgggtggttgaggcgg  
tgaggagggcgaacgggggagggccaggcgggcccggcgggcggtggcgggggaggggcgcg  
cggcgggaaggagcgggtggcccttcgggcgccttgcgcgccctcggagccgcgcctcg  
tgggcgcgcggcggaaaaactggagcggcagaatgaatggggctctcttttcgggcgcgc  
gcggcggagctttctcgggcctgggggtcgggactggctgatcgcccgacgtccgggcct
```

El cambio se encuentra en la posición justo después del exón 1 (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

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
421		2	+	1.00	AGAAAGACAA	^	GTGAGTGGGA	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'
233		1	+	0.27	CCACGCGCAG	^	GCAGCACCTC	
313		0	+	0.19	GTCTCCCGAG	^	GGCGTCGGCA	
324		2	+	0.19	GCGTCGGCAG	^	GGTCGGCGGC	
342		2	+	0.07	GCGTCGGCAG	^	CAGTGTGAC	
345		2	+	0.07	TCGGCAGCAG	^	TGTCGACGGC	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
620		101		1	-	0.53	CCCGCTCCAG	^	TTTTCCGCCG	
400		321		0	-	0.32	GGTGCCGAAG	^	ATGGAGGCCA	
394		327		0	-	0.32	GAAGATGGAG	^	GCCAGATACT	
389		332		2	-	0.25	TGGAGGCCAG	^	ATACTCCGCC	
119		602		0	-	0.18	CCGGCCCTAG	^	CTCTCTCGCC	

Donor splice sites, direct strand

No donor site predictions above threshold.

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'
233		1	+	0.27	CCACGCGCAG	^	GCAGCACCTC	
313		0	+	0.19	GTCTCCCGAG	^	GGCGTCGGCA	
324		2	+	0.18	GCGTCGGCAG	^	GGTCGGCGGC	

Acceptor splice sites, complement strand

pos	3'→5'	pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
620		101		1	-	0.53	CCCGCTCCAG	^	TTTTCCGCCG	
400		321		0	-	0.25	GGTGCCGAAG	^	ATGGAGGCCA	
394		327		0	-	0.32	GAAGATGGAG	^	GCCAGATACT	
389		332		2	-	0.25	TGGAGGCCAG	^	ATACTCCGCC	
119		602		0	-	0.18	CCGGCCCTAG	^	CTCTCTCGCC	

No se predice ningún sitio *donor* para la secuencia mutante, por lo que se pierde el *donor* de la secuencia wt, lo que provocará probablemente el *skipping* del exón 1.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.15.54.30368.0 :

Start	End	Score	Exon	Intron
414	428	0.97	aagacaa	gt gagtgg

Donor site predictions for 85.53.15.54.30389.0 :

Start	End	Score	Exon	Intron
-------	-----	-------	------	--------

Acceptor site predictions for 85.53.15.54.30368.0 : Acceptor site predictions for 85.53.15.54.30389.0 :


Start	End	Score	Intron	Exon	Start	End	Score	Intron	Exon
-------	-----	-------	--------	------	-------	-----	-------	--------	------

Pérdida sitio *donor* en la secuencia mutante. Este coincide con el sitio *donor* del exón 1, por lo que estará afectando al *splicing*, probablemente perdiéndose dicho exón ya que no hay otro *donor* que se pueda usar.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
gacaa(g/a)tgagt	aagtga	aaatga	28814	66%

Human Splicing Finder

 Broken WT Donor Site	Alteration of the WT Donor site, most probably affecting splicing		
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr21:43107453	- REF : CAA G TGAGT - ALT : CAA A TGAGT	87.07 > 59.93 => -31.17%
MaxEnt Donor site	chr21:43107453	- REF : CAA G TGAGT - ALT : CAA A TGAGT	9.1 > 0.91 => -90%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr			
wt	18	323	gcctgactg		3.14152760285	0.380503144654	137	18	27			-6.6656164
wt	18	270	gggtgacgt		1.87297258093	0.37358490566	84	18	27			-3.809739
wt	18	148	aagtgagt		-0.878641650193	0.342657342657	8	9	14			-0.20754832
wt	18	97	ggttgaggc		-1.09725035719	0.315217391304	92	0	0			-5.7494871
seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr			
mut	18	323	gcctgactg		3.14152760285	0.380503144654	137	18	27			-6.6656164
mut	18	270	gggtgacgt		1.87297258093	0.37358490566	84	18	27			-3.809739
mut	18	148	aaatgagt		-0.985617479759	0.342657342657	8	9	14			-0.2494344
mut	18	97	ggttgaggc		-1.09725035719	0.315217391304	92	0	0			-5.7494871

El único cambio es el BP en 148 que pasa de tener una g en la tercera posición a una a (la mutación). Siguen siendo puntuaciones negativas, por lo que no se tiene en cuenta.

Variant Effect Predictor tool

ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	splice_donor_variant	U2AF1	ENSG00000160201	Transcript	ENST00000291552.9	protein_coding	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	splice_donor_variant	U2AF1	ENSG00000160201	Transcript	ENST00000380276.6	protein_coding	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	upstream_gene_variant	U2AF1	ENSG00000160201	Transcript	ENST00000459639.5	protein_coding	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	upstream_gene_variant	U2AF1	ENSG00000160201	Transcript	ENST00000463599.5	retained_intron	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	splice_donor_variant, NMD_transcript_variant	U2AF1	ENSG00000160201	Transcript	ENST00000464750.5	nonsense_mediated_decay	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	splice_donor_variant, non_coding_transcript_variant	U2AF1	ENSG00000160201	Transcript	ENST00000475639.5	retained_intron	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	splice_donor_variant, NMD_transcript_variant	U2AF1	ENSG00000160201	Transcript	ENST00000486519.5	nonsense_mediated_decay	COSV52347342
ENST00000291552.8:c.44+1G>A	21:43107450-43107450	T	splice_donor_variant, non_coding_transcript_variant	U2AF1	ENSG00000160201	Transcript	ENST00000496462.1	retained_intron	COSV52347342

Se trata de una variante que está afectando al sitio *donor* normal del *splicing*.

ESEfinder

Se obtienen dos resultados (406 y 410) donde se encuentra la posición de interés (g detrás de las mayúsculas que indican el exón) con puntuaciones positivas, para las matrices de 5'SS:

406 (-315)	CACCGAGAAAGACAAgtgagtgaggagcccc	11.30070	406 (-315)	CACCGAGAAAGACAAgtgagtgaggagcccc	-15.39060	406 (-315)	CACCGAGAAAGACAAgtgagtgaggagcccc	11.79020	406 (-315)	CACCGAGAAAGACAAgtgagtgaggagcccc	-19.28390
407 (-314)	ACCGAGAAAGACAAgtgagtgaggagcccc	-25.02360	407 (-314)	ACCGAGAAAGACAAgtgagtgaggagcccc	-9.11340	407 (-314)	ACCGAGAAAGACAAgtgagtgaggagcccc	-23.28730	407 (-314)	ACCGAGAAAGACAAgtgagtgaggagcccc	-9.97280
408 (-313)	CCGAGAAAGACAAgtgagtgaggagcccc	-11.80160	408 (-313)	CCGAGAAAGACAAgtgagtgaggagcccc	-33.58010	408 (-313)	CCGAGAAAGACAAgtgagtgaggagcccc	-7.70070	408 (-313)	CCGAGAAAGACAAgtgagtgaggagcccc	-32.95830
409 (-312)	CGAGAAAGACAAgtgagtgaggagcccccg	-23.25620	409 (-312)	CGAGAAAGACAAgtgagtgaggagcccccg	-24.09050	409 (-312)	CGAGAAAGACAAgtgagtgaggagcccccg	-21.08140	409 (-312)	CGAGAAAGACAAgtgagtgaggagcccccg	-26.13480
410 (-311)	GAGAAAGACAAgtgagtgaggagcccccg	3.91320	410 (-311)	GAGAAAGACAAgtgagtgaggagcccccg	-36.01080	410 (-311)	GAGAAAGACAAgtgagtgaggagcccccg	4.35770	410 (-311)	GAGAAAGACAAgtgagtgaggagcccccg	-38.00880

Si buscamos la resultado equivalente en la secuencia mutante, se observa que las puntuaciones para las matrices 5'SS se reducen considerablemente para 406 y se mantienen más o menos iguales para 410:

406 (-315)	CACCGAGAAAGACAAatgagtgggagcccc	0.31820	406 (-315)	CACCGAGAAAGACAAatgagtgggagcccc	-16.72060	406 (-315)	CACCGAGAAAGACAAatgagtgggagcccc	0.66190	406 (-315)	CACCGAGAAAGACAAatgagtgggagcccc	-20.53430
407 (-314)	ACCGAGAAAGACAAatgagtgggagcccc	-28.41100	407 (-314)	ACCGAGAAAGACAAatgagtgggagcccc	-23.09830	407 (-314)	ACCGAGAAAGACAAatgagtgggagcccc	-26.57620	407 (-314)	ACCGAGAAAGACAAatgagtgggagcccc	-27.32080
408 (-313)	CCGAGAAAGACAAatgagtgggagcccc	-9.49610	408 (-313)	CCGAGAAAGACAAatgagtgggagcccc	-17.00080	408 (-313)	CCGAGAAAGACAAatgagtgggagcccc	-5.33810	408 (-313)	CCGAGAAAGACAAatgagtgggagcccc	-18.42100
409 (-312)	CGAGAAAGACAAatgagtgggagcccccg	-22.70800	409 (-312)	CGAGAAAGACAAatgagtgggagcccccg	-19.95110	409 (-312)	CGAGAAAGACAAatgagtgggagcccccg	-20.44060	409 (-312)	CGAGAAAGACAAatgagtgggagcccccg	-21.52650
410 (-311)	GAGAAAGACAAatgagtgggagcccccg	4.09900	410 (-311)	GAGAAAGACAAatgagtgggagcccccg	-36.08200	410 (-311)	GAGAAAGACAAatgagtgggagcccccg	4.49750	410 (-311)	GAGAAAGACAAatgagtgggagcccccg	-38.00740

Por lo tanto, lo que nos indican estos resultados es que se está debilitando un sitio *donor* (406), que puede estar afectando al *splicing* normal.