

## Ejemplo comparación de resultados predictores in silico

Cambio de estudio GIT1 c.405+5G>C (chr17:29582693 G/C, COSV56608869 o NM\_014030.4:c.405+5G>C)

Exón 4 e intrones adyacentes:

```
gtgagtgggctggagggtgtgtgagagactgcgcagtcgsgacagggccctggga
ggtggcaaggtgaccaccctctcccaccatgctcccattcctcctgtgtgcccattgca
g
CCCCATCAAGTCAGAGTTCATCAGGGCCAAGTACCAGATGCTGGCATTTGTGCACAAGCT
TCCCTGCCGGGACGATGATGGAGTCACCGCCAAGACCTCAGCAAG
gtttgagggccctacaacagatgggtggccccctcctctctctccatccaagaactctc
tgaaggcctctcctttgtccccactgagccctgggcccacaaggcagccaggtcagggtta
acagtagaacaggtggagagagcagattgtttgggtcaggccgtgtgtggctgaggctccc
cagcaaccaagtagacacacaaggctagaagcccaagctgcactcagaagctgggaaa
acctgggggttagttgcttagccactcatagcctcagtttactcatdgtataaatgggaaa
gaatatacagcatctacttcataaatagtatctaagcagggaacacagcacaagtggttag
atggtgactgacggcagcgtttattaattggcattaacccctagcgggaagcctctgtcc
tcccgacatagctctgttttgggcaggcctccttggaagctgctaggtgtccacaggggcagc
caggtgcagcttggtgggtgggaggtgcgcctcacgcacatgcgtattcactgcacact
ctgctcctcttag
```

El cambio se encuentra en quinta posición justo después del exón 4 (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'	exon	intron	3'
223		1	+	0.06	AAAGACCTCA	GCAAGGTTTG		
228		0	+	0.91	CCTCAGCAAG	GTTTGAGGGG	H	
339		1	+	0.39	AGGCAGCCAG	GTCAGGGTAA		

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
155	626	0	-	0.61	CCAGCATCTG	GTACTTGGCC		

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
121		2	+	0.77	CCCCATGCAG	CCCCATCAAG		
131		0	+	0.42	CCCCATCAAG	TCAGAGTTCA		
135		1	+	0.34	ATCAAGTCAG	AGTTCATCAG		
137		0	+	0.32	CAAGTCAGAG	TTCATCAGGG		
145		2	+	0.19	AGTTCATCAG	GGCCAAGTAC		
152		0	+	0.07	CAGGGCCAAG	TACCAGATGC		

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
559	222	1	-	0.17	CCCTGCTTAG	ATACTATTTA		
513	268	2	-	0.43	CATTTTACAG	ATGAGTAAAC		
240	541	0	-	0.34	TCTGTTGTAG	GGCCCCTCAA		
219	562	0	-	0.19	CCTTGCTGAG	GTCTTTGGCG		
185	596	0	-	0.43	GTCCCGGCAG	GGAAGCTTGT		
55	726	2	-	0.00	CACCTCCCAG	GGCCCTGTCC		

### Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
228		0	+	0.55	CCTCAGCAAG	GTTTGAGGGG		
339		1	+	0.39	AGGCAGCCAG	GTCAGGGTAA		

### Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
155	626	0	-	0.55	CCAGCATCTG	GTACTTGGCC		

### Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
121		2	+	0.77	CCCCATGCAG	CCCCATCAAG		
131		0	+	0.43	CCCCATCAAG	TCAGAGTTCA		
135		1	+	0.34	ATCAAGTCAG	AGTTCATCAG		
137		0	+	0.32	CAAGTCAGAG	TTCATCAGGG		
145		2	+	0.19	AGTTCATCAG	GGCCAAGTAC		
152		0	+	0.07	CAGGGCCAAG	TACCAGATGC		

### Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
559	222	1	-	0.17	CCCTGCTTAG	ATACTATTTA		
513	268	2	-	0.43	CATTTTACAG	ATGAGTAAAC		
240	541	0	-	0.17	TCTGTTGTAG	GGCCCCTCAA		
219	562	0	-	0.18	CCTTGCTGAG	GTCTTTGGCG		
185	596	0	-	0.54	GTCCCGGCAG	GGAAGCTTGT		
55	726	2	-	0.00	CACCTCCCAG	GGCCCTGTCC		

Pérdida sitio *donor* (en rojo) en la secuencia mutante. Esta región se encuentra en el interior del exón, por lo que no se emplea en el *splicing* normal del mRNA. La mutación no tendrá efecto en el *splicing*.

### Splice Site Prediction by Neural Network (NNSplice)

#### Donor site predictions for 89.130.114.18.17096.0 :

Start	End	Score	Exon	Intron
63	77	0.41	tggcaag	gtgaccac
221	235	0.62	cagcaag	gtttgagg
332	346	0.80	cagccag	gtcagggt

#### Donor site predictions for 89.130.114.18.17110.0 :

Start	End	Score	Exon	Intron
63	77	0.41	tggcaag	gtgaccac
332	346	0.80	cagccag	gtcagggt

#### Acceptor site predictions for 89.130.114.18.17096.0 :

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

#### Acceptor site predictions for 89.130.114.18.17110.0 :

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

Pérdida de un sitio *donor* (en rojo) en la secuencia mutante. Este en concreto es el sitio *donor* del exón 4, pues se encuentra en el final del exón y principio del intrón 4. Por lo tanto, si se pierde un sitio *donor* puede producirse o el *skipping* del exón o la inclusión de un exón críptico.

**MaxEntScan** (solo realizamos el análisis del 5'SS porque está por delante del exón y, si afecta a un sitio de *splicing*, será de este tipo)

MAXENT: -9.19    MDD: -6.05    MM: -8.99    WMM: -9.68

>sec

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ggttt(g/c)agggg	gagggg	cagggg	26469	52%

Broken WT Donor Site		Alteration of the WT Donor site, most probably affecting splicing	
Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr17:29582700	- REF : AAGGTTTGA - ALT : AAGGTTTCA	83.8 > 73.79 => -11.95%
MaxEnt Donor site	chr17:29582700	- REF : AAGGTTTGA - ALT : AAGGTTTCA	5.13 > -2.05 => -139.96%

 New Acceptor splice site	Activation of a cryptic Acceptor site. Potential alteration of splicing		
Algorithm/Matix	position	sequences	variation
HSF Acceptor site (matrix AG)	chr17:29582702	- REF : GCAAGGTTT <b>G</b> AGGG - ALT : GCAAGGTTT <b>C</b> AGGG	70.25 > 77.89 => 10.88%

seq_id	agez	ss_dist	bp_eq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr								
wt	15	402	aggtagacca		1.42749393983	0.5163727597	1	15	24	1.2877645							
wt	15	346	ccatcaagt		-2.43633918142	0.478005865103	51	10	19	-3.4489955							
wt	15	341	aagtcagag		-2.75129231578	0.479166666667	46	10	19	-3.255448							
wt	15	334	agttcatca		-1.31492693658	0.480243161094	39	10	19	-2.2496068							
wt	15	331	tcatacagg		-3.31218636923	0.478527607362	36	10	19	-2.8422878							
wt	15	276	cgatgatgg		0.809732724934	0.479784797048	58	24	47	-2.3597008							
wt	15	268	gagtcaccg		0.722488771671	0.48288973384	50	24	47	-1.8864458							
wt	15	253	acctcagca		-0.287534839085	0.479838709677	35	24	47	-1.3334283							
wt	15	242	gtttgaggg		-2.00504619751	0.485232067511	24	24	47	-1.3078917							
wt	15	185	ctctgaagg		0.726595176089	0.438888888889	4	19	37	0.91951544							
wt	15	161	cactgacgc		1.77767596887	0.397435897436	132	9	12	-7.0174061							
wt	15	133	aggtagcagg		-2.8971086997	0.3828125	104	9	12	-7.0801752							
wt	15	127	gggtaacag		1.36380865899	0.3934462262951	98	9	12	-5.0286015							
wt	15	91	gggtcagcg		-2.08459464497	0.4302325814	62	9	12	-4.0881928							
wt	15	75	gggtcagcg		0.352138110667	0.428651428571	46	9	12	-2.12186							
wt	15	20	cactcagaa		-0.414550395741	0.266666666667	15	0	0	-0.6238874							

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr				
mut	15	402	agggtgacca		1.42749393983	0.518991687657	1	15	24	1.288578			
mut	15	346	ccatcaagt		-2.43633918142	0.480938416422	51	10	19	-3.4480484			
mut	15	341	aagtcagag		-2.75129231578	0.482142857143	46	10	19	-3.2544868			
mut	15	334	agttcagca		-1.31492603658	0.483282674772	39	10	19	-2.2486251			
mut	15	331	tcatacaggg		-3.31218636923	0.481595092025	36	10	19	-2.8412971			
mut	15	276	cgatgcattg		0.809732724934	0.483394833948	58	24	47	-2.3585091			
mut	15	268	gagtcaccg		0.722488771671	0.486692015209	50	24	47	-1.8852177			
mut	15	253	acctcagca		-0.8754839985	0.483879967742	35	24	47	-1.332126			
mut	15	242	gtttcagg		-2.38305715253	0.485232067511	24	24	47	-1.4559089			
mut	15	185	ctctgaagg		0.726595176089	0.438888888889	4	19	37	0.91951544			
mut	15	161	cactgagcc		1.77767596887	0.397435897436	132	9	12	-7.6174061			
mut	15	133	agggtcagg		-2.8971086997	0.3828125	104	9	12	-7.0801752			
mut	15	127	gggttaacag		1.36380865899	0.393442622951	98	9	12	-5.0286015			
mut	15	91	gggtcaggc		-2.08459464497	0.43023255814	62	9	12	-4.0881928			
mut	15	75	ggctcaggc		0.352138110667	0.428571428571	46	9	12	-2.12186			
mut	15	20	cactcagaa		-0.414550395741	0.266666666667	15	0	12	-0.62388748			

La diferencia entre ambas es el BP donde se encuentra la mutación de interés en la última posición (en amarillo). El BP tiene puntuación menos negativa en la secuencia mutante, pero siguen siendo números muy bajos, por lo que es probable que no esté afectando al *splicing*.

Se trata de una variante que está afectando al sitio de *splicing* (7/13 resultados), por lo que va a provocar que se altere el *splicing* normal, probablemente en la región *donor* dada su proximidad.

## Variant Effect Predictor tool

ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000225394.8</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000335356.7</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000394869.7</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	upstream_gene_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000473217.5</a>	retained_intron	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	downstream_gene_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000577466.2</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant, non_coding_transcript_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000579536.1</a>	retained_intron	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000579937.5</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000581348.5</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	intron_variant, non_coding_transcript_variant	ABHD15-AS1	<a href="#">ENSG00000264031</a>	Transcript	<a href="#">ENST00000581474.1</a>	lncRNA	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	upstream_gene_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000581925.1</a>	retained_intron	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	splice_region_variant, intron_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000583413.4</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	upstream_gene_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000585148.1</a>	protein_coding	-	-	<a href="#">COSV56608869</a>
ENST00000581348.5:c.405+5G>C	<a href="#">17-29582693-29582693</a>	G	upstream_gene_variant	GIT1	<a href="#">ENSG00000108262</a>	Transcript	<a href="#">ENST00000586574.5</a>	retained_intron	-	-	<a href="#">COSV56608869</a>

## ESEfinder

Solo se obtiene un resultado positivo donde se encuentre la posición de estudio (segunda g después de las mayúsculas del exón) para las matrices 5'SS:

213 (-568)	AAAGACCTCAGCAAGgtttgaggggccccta	5.86180	213 (-568)	AAAGACCTCAGCAAGgtttgaggggccccta	-3.22350	213 (-568)	AAAGACCTCAGCAAGgtttgaggggccccta	5.89020	213 (-568)	AAAGACCTCAGCAAGgtttgaggggccccta	-3.95580
---------------	---------------------------------	---------	---------------	---------------------------------	----------	---------------	---------------------------------	---------	---------------	---------------------------------	----------

Si buscamos el resultado equivalente para la secuencia mutante, se observa que las puntuaciones son menores para esas matrices:

213 (-568)	AAAGACCTCAGCAAGgtttcaggggccccta	2.05890	213 (-568)	AAAGACCTCAGCAAGgtttcaggggccccta	-2.85680	213 (-568)	AAAGACCTCAGCAAGgtttcaggggccccta	2.11040	213 (-568)	AAAGACCTCAGCAAGgtttcaggggccccta	-3.63990
---------------	---------------------------------	---------	---------------	---------------------------------	----------	---------------	---------------------------------	---------	---------------	---------------------------------	----------

Por lo tanto, se está debilitando el sitio 5'SS, provocando una posible alteración del *splicing*.