

Ejemplo comparación de resultados predictores in silico

Cambio de estudio COL5A1 c.925-2A>G (chr9: 134730234 A/G, CS113078 o NM_000093.3_c.925-2A>G)

Exón 7 e intrones adyacentes:

```
ttaaatactgagatatgattcacccittttctgggtgtgtaacbggggtgggtttgagtcacat
tctgagcacgcttagtaactgggggcttgacagctgggctgctctcttccacattgctctg
cccccaagcccatggatgggggcccctgcacccaagaggctctctgggctcttgac
aggcctgggctccaggcggttgccactgagatgcctgcacccggacatggcggaagtcacca
gacctgggcaccactgcggcctccgcctgactccagctgtctctgtccttggctcccag
GAGCTGACCCCGACCCACCGGAAGCTGCTCCCATGCCCTGAAACCAGTGAAGGGGCTGGG
AAGGAAGAGGACGTCGGCATCGGGGACTATGACTACGTGCCAGTGAGGACTACTACCG
CCCTCACCTATCATGACCTCACCTATGGCGAGGGGGAGGAGAACCCGACCAGCCACA
GACCCAGGCGCTGGGGCCGAAATTCCACACAGCACCGCCGACACCTCCAACTCCTCCAAT
gtaatttctttccttccattgttttggtctggggcagtgggccaagggcagggctggg
gccacaatgctgagctccttcttactccagttctcacccttggtgtcctcgggtggcccc
tggtttccacacacccctggccctgggdcctttgcagctgggtggcgtaatacttccag
gggaggccaattgctgaacaagtgttgaatdcctgggtttgctgagttcatctcaatgttgt
caattttttgctgggaaatgggtagaaatgaccctagcccttcgggtcttgaggattca
```

Se ha descrito que este cambio causa la producción de varios transcritos: (1) donde se produce el *skipping* del exón 7 y 6 y (2) con el *skipping* del exón 7. El *spliceosome* utiliza el sitio *donor* del exón 5 y el *acceptor* del exón 8 llevándose a cabo el *skipping* de ambos exones¹.

El cambio se encuentra en segunda posición justo antes del exón 7 (la **a** en color rojo).

Se va a obtener los resultados que produce analizar esta variable con los diferentes predictores y ver cuál de ellos es más preciso.

1. Symoens, S., Malfait, F., Vlummen, P., Hermanns-Lê, T., Syx, D., & De Paepe, A. (2011). A novel splice variant in the N-propeptide of COL5A1 causes an EDS phenotype with severe kyphoscoliosis and eye involvement. PLoS one, 6(5), e20121. <https://doi.org/10.1371/journal.pone.0020121>

NetGene2

The sequence: wt has the following composition:

Length: 840 nucleotides.

18.2% A, 31.4% C, 27.6% G, 22.7% T, 0.0% X, 59.0% G+C

Donor splice sites, direct strand

No donor site predictions above threshold.

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
638	203	0	-	0.67	GGACACCAAG	GTGAGAACTG		
443	398	1	-	0.60	CTCGCCATAG	GTGAGGTCAT		
427	414	2	-	0.81	TCATCATACG	GTGAGGGCGT		

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
182	2	+	0.19	CTCTTGACAG	GCCTGGGCTC		
195	0	+	0.14	TGGGCTCCAG	GCGTTGCCAC		
209	2	+	0.07	TGCCACTGAG	ATGCCTGCAC		
277	1	+	0.14	CTGACTCCAG	CTGTCTCTGT		
300	0	+	0.93	TGGCTCCCAG	GAGCTGACCC		
303	0	+	0.17	CTCCCAGGAG	CTGACCCCGA		
631	0	+	0.27	CTTACTCCAG	TTCTCACCTT		

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
596	245	1	-	0.14	GTGGCCCCAG	CCCTGGCCCT		
570	271	1	-	0.34	ACTGCCCCAG	ACCAAACCAA		
553	288	0	-	0.17	CAATGGGAAG	GAAAGAAATT		
444	397	1	-	0.25	CTCGCCATAG	GTGAGGTCAT		
356	485	2	-	0.33	TCCTTCCCAG	CCCCTTCACT		

The sequence: mut has the following composition:

Length: 840 nucleotides.

18.1% A, 31.4% C, 27.7% G, 22.7% T, 0.0% X, 59.2% G+C

Donor splice sites, direct strand

No donor site predictions above threshold.

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
638	203	0	-	0.67	GGACACCAAG	GTGAGAACTG		
443	398	1	-	0.61	CTCGCCATAG	GTGAGGTCAT		
427	414	2	-	0.81	TCATCATACG	GTGAGGGCGT		

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
182	2	+	0.18	CTCTTGACAG	GCCTGGGCTC		
195	0	+	0.14	TGGGCTCCAG	GCGTTGCCAC		
209	2	+	0.07	TGCCACTGAG	ATGCCTGCAC		
235	1	+	0.07	ATGCGGCAAG	TCCCAGACCT		
277	1	+	0.14	CTGACTCCAG	CTGTCTCTGT		
303	0	+	0.07	CTCCGGGAG	CTGACCCCGA		
631	0	+	0.27	CTTACTCCAG	TTCTCACCTT		

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
596	245	1	-	0.14	GTGGCCCCAG	CCCTGGCCCT		
570	271	1	-	0.34	ACTGCCCCAG	ACCAAACCAA		
553	288	0	-	0.17	CAATGGGAAG	GAAAGAAATT		
444	397	1	-	0.25	CTCGCCATAG	GTGAGGTCAT		
356	485	2	-	0.33	TCCTTCCCAG	CCCCTTCACT		

Se predicen los mismos sitios de *splicing*, luego esta mutación no tendrá efecto en el *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.81.70.8121.0 :

Start	End	Score	Exon	Intron
26	40	0.69	ttttctg	gtgtgtaa
38	52	0.42	taactgg	gtgggttt

Donor site predictions for 85.53.81.70.8241.0 :

Start	End	Score	Exon	Intron
26	40	0.69	ttttctg	gtgtgtaa
38	52	0.42	taactgg	gtgggttt

Acceptor site predictions for 85.53.81.70.8121.0 :

Start	End	Score	Intron	Exon
162	202	0.81	gggtctctgggcctcttgac	aggcctgggctccaggcggtgc
280	320	0.66	gtctctgtccttggtctccc	aggagctgaccccgacccccac
611	651	0.92	tgagctcccttcttactcc	agttctcaccttggtgtcctcg
677	717	0.74	ctggccctgggcccctttgc	agctgggtggcgtaatacttcc

Acceptor site predictions for 85.53.81.70.8241.0 :

Start	End	Score	Intron	Exon
162	202	0.81	gggtctctgggcctcttgac	aggcctgggctccaggcggtgc
611	651	0.92	tgagctcccttcttactcc	agttctcaccttggtgtcctcg
677	717	0.74	ctggccctgggcccctttgc	agctgggtggcgtaatacttcc

Desaparece un sitio *aceptor* en la predicción para la secuencia mutante (derecha). Este sitio (en rojo) incluye a la posición de estudio. Por lo tanto, el *spliceosome* no podrá utilizar el sitio aceptor del exón 7, produciendo el *skipping* de este exón o, en el caso de que se utilice otro sitio de los predichos, podría perderse parte del exón o ganar parte del intrón dentro del mRNA.

GENSCAN → no da resultados para este cambio

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

NO EXONS/GENES PREDICTED IN SEQUENCE

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

NO EXONS/GENES PREDICTED IN SEQUENCE

MaxEntScan (solo se estudia 3' por la proximidad con el cambio)

>sec
ctctgtccttggctcccagGAGCTGACCCCGACCCACGG MAXENT: -2.74 MM: -3.42 WMM: -0.94

La posición no es muy susceptible a cambios, dado que no se encuentra en un sitio débil de *splicing* (valores negativos para los diferentes indicadores).

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ctccc(a/g)ggagc	cccagg	cccggg	31493	83%


En el análisis de la región adyacente al cambio, se obtiene una puntuación elevada (83%) para el cambio A>G, por lo que puede estar afectando al *splicing*.

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

Type	↑↓		Interpretation	↑↓	
 Broken WT Acceptor Site			Alteration of the WT Acceptor site, most probably affecting splicing		
Algorithm/Matix	position	sequences		variation	
MaxEnt Acceptor site	chr9:134730216	- REF : TCTCTGTCCTTGGCTCCCAGGAG - ALT : TCTCTGTCCTTGGCTCCC G GGAG		8.95 > 0.99 => -88.94%	
HSF Acceptor site (matrix AG)	chr9:134730224	- REF : CTTGGCTCCCAGGA - ALT : CTTGGCTCCC G GGA		86.61 > 58.74 => -32.18%	

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_of	ppt_len	ppt_scr	svm_scr		seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr	
wt	19	472	tcttgacag		0.666890937047	0.528907922912	32	11	12	-1.0800372	mut	19	472	tcttgacag		0.666890937047	0.528907922912	32	11	12	-1.0800372
wt	19	443	cactgagat		0.546149470404	0.529680365297	3	11	12	0.70858762	mut	19	443	cactgagat		0.546149470404	0.529680365297	3	11	12	0.70858762
wt	19	380	ccctgactc		3.67444049564	0.525333333333	7	21	34	1.8838127	mut	19	380	ccctgactc		3.67444049564	0.525333333333	7	21	34	1.8838127
wt	19	344	agctgaccc		3.27565441959	0.513274336283	1	13	14	1.9172474	mut	19	344	agctgaccc		3.27565441959	0.513274336283	1	13	14	1.9172474
wt	19	310	gcctgaaac		1.1478208853	0.501639344262	78	10	13	-3.8029443	mut	19	310	gcctgaaac		1.1478208853	0.501639344262	78	10	13	-3.8029443
wt	19	301	cagtgaagg		-0.559704308555	0.506756756757	69	10	13	-3.9001834	mut	19	301	cagtgaagg		-0.559704308555	0.506756756757	69	10	13	-3.9001834
wt	19	259	ctatgacta		0.18020398317	0.547244094488	27	10	13	-0.93886864	mut	19	259	ctatgacta		0.18020398317	0.547244094488	27	10	13	-0.93886864
wt	19	244	cagtgagga		-1.59392083786	0.548117154812	12	10	13	-0.68376556	mut	19	244	cagtgagga		-1.59392083786	0.548117154812	12	10	13	-0.68376556
wt	19	225	ccctcaccg		3.28702863303	0.545454545455	12	10	15	1.2451286	mut	19	225	ccctcaccg		3.28702863303	0.545454545455	12	10	15	1.2451286
wt	19	217	gtatgatga		-0.601645845772	0.547169811321	4	10	15	0.22946929	mut	19	217	gtatgatga		-0.601645845772	0.547169811321	4	10	15	0.22946929
wt	19	214	tgatgacct		1.68733892888	0.55023923445	1	10	15	1.3166013	mut	19	214	tgatgacct		1.68733892888	0.55023923445	1	10	15	1.3166013
wt	19	209	acctcacct		1.74333388753	0.544117647059	80	16	23	-3.5894917	mut	19	209	acctcacct		1.74333388753	0.544117647059	80	16	23	-3.5894917
wt	19	107	atgtaattt		-0.618901187121	0.598039215686	1	17	40	0.66193341	mut	19	107	atgtaattt		-0.618901187121	0.598039215686	1	17	40	0.66193341
wt	19	38	tgctgagct		1.65422834964	0.757575757576	2	27	49	1.624041	mut	19	38	tgctgagct		1.65422834964	0.757575757576	2	27	49	1.624041
wt	19	26	ttcttactc		0.712943788249	0.714285714286	1	16	26	1.0905361	mut	19	26	ttcttactc		0.712943788249	0.714285714286	1	16	26	1.0905361

No hay cambios entre la predicción para la secuencia WT con respecto a la mutante, por lo que no debe estar afectando al *splicing*.

IntSplice

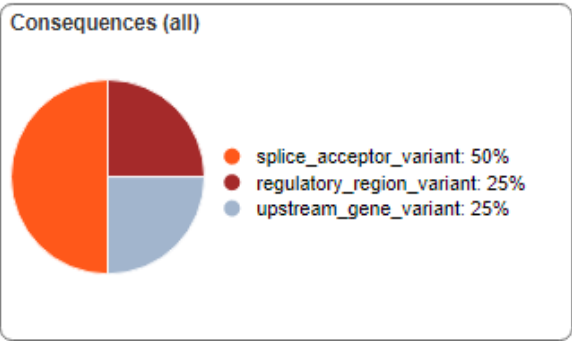
SNV at chr9:134730234 can't be predicted by IntSplice.

Prediction shows either Abnormal or Normal.

Prediction	Genomic Mutation	Ensembl 64 Transcript ID and Exon No.
------------	------------------	---------------------------------------

Variant Effect Predictor tool

Category	Count
Variants processed	1
Variants filtered out	0
Novel / existing variants	0 (0.0) / 1 (100.0)
Overlapped genes	1
Overlapped transcripts	3
Overlapped regulatory features	1



Se trata de una variante que está afectando al sitio de *splicing*, por lo que va a provocar que se altere el *splicing* normal.

Uploaded variant	Location	Allele	Consequence	Symbol	Gene	Feature type
ENST00000371817.8:c.925-2A>G	9:134730234-134730234	G	splice_acceptor_variant	COL5A1	ENSG00000130635	Transcript
ENST00000371817.8:c.925-2A>G	9:134730234-134730234	G	upstream_gene_variant	COL5A1	ENSG00000130635	Transcript
ENST00000371817.8:c.925-2A>G	9:134730234-134730234	G	splice_acceptor_variant	COL5A1	ENSG00000130635	Transcript
ENST00000371817.8:c.925-2A>G	9:134730234-134730234	G	regulatory_region_variant	-	-	RegulatoryFeature

ESEfinder

286 (-555)	gtccttggtcccagGAGCTGACCCCGACC	-10.03210	286 (-555)	gtccttggtcccagGAGCTGACCCCGACC	11.22410	286 (-555)	gtccttggtcccagGAGCTGACCCCGACC	-6.25290	286 (-555)	gtccttggtcccagGAGCTGACCCCGACC	11.21450
287 (-554)	tccttggtcccagGAGCTGACCCCGACCC	-24.11670	287 (-554)	tccttggtcccagGAGCTGACCCCGACCC	-9.63240	287 (-554)	tccttggtcccagGAGCTGACCCCGACCC	-22.18370	287 (-554)	tccttggtcccagGAGCTGACCCCGACCC	-7.53370
288 (-553)	ccttggtcccagGAGCTGACCCCGACCCC	-10.70730	288 (-553)	ccttggtcccagGAGCTGACCCCGACCCC	-26.91600	288 (-553)	ccttggtcccagGAGCTGACCCCGACCCC	-10.91210	288 (-553)	ccttggtcccagGAGCTGACCCCGACCCC	-28.76690
289 (-552)	cttggtcccagGAGCTGACCCCGACCCC	-3.39060	289 (-552)	cttggtcccagGAGCTGACCCCGACCCC	0.52340	289 (-552)	cttggtcccagGAGCTGACCCCGACCCC	-3.12660	289 (-552)	cttggtcccagGAGCTGACCCCGACCCC	-0.18630

La posición de interés no aparece como el mejor resultado del buscador, pero, cuando se buscan todas las secuencias, se obtiene una puntuación positiva para las matrices de 3'SS en dos resultados (286 y 289) por lo que esta deberá ser la región que produzca este sitio de *splicing*. Si se hace la búsqueda para la secuencia mutada, la posición de interés sale solo en un resultado (289) en una de las dos matrices de 3'SS con una puntuación muy baja (0.37830). Por lo tanto, se está perdiendo el sitio *acceptor* (el resultado 286), afectando el *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	5	4	23	275.2194	5	-9.5697	23	25	119	1486.1293	92	131.2235	37	259	0.14
mut	0	5	4	23	275.2194	5	-9.5697	23	25	119	1486.1293	92	131.2235	37	259	0.14

Both alleles have a comparable chance of exon skipping.

HOT-SKIP

HOT-SKIP - Results for submitted sequences

```
>wt
ttaaatactgagatatgattcacccttttctgggtgtgtaactgggtgggtttgagtccattctgagcacgcttagtaact
gggggcttgagctgggctgcttcttcccacttgctctgcccccaagcccatggatgggggaggccctgcacccaag
aggtctctgggcctcttgacaggcctgggctccaggcgttgccactgagatgctgcacccggacatgcggcaagtcca
gacctggcaccactgccggcctccgccctgactccagctgtctctgtccttggctcccagGAGCTGACCCGACCCAC
GGAAGCTGCTCCCATGCCTGAAACCAGTGAGGGGCTGGGAAAGGAAGAGGACGTCGGCATCGGGGACTATGACTACGTGC
CCAGTGAGGACTACTACACGCCCTACCGTATGATGACCTACCTATGGCGAGGGGGAGGAGAACCCGACCAGCCACA
GACCCAGGCGCTGGGGCCGAAATTCACACAGCACCCGCCGACACCTCCAACCTCCTCAATgtaatttcttcttccat
tggtttggtctggggcagtgggccaaggccaggctggggccacaatgctgagctcccttcttactccagttctcacct
tgggtgcctcgggtggccctgtgttccaccacccctggccctgggccccttgcagctgggtggcgtaatacttcccag
gggaggccattgctgaacaagtttggaatccctggtttgctgagttcatctcaatgttgtaatttttgcctggaaaat
gggtagaatgacctagcccttccggtctttgaggattca
```

Mutation(s) E+56C>G, E+158A>G and E+56C>T have the highest probability of exon skipping.

No marca la mutación de interés.

Como este predictor se emplea para regiones exónicas, su información no es útil para una variable intrónica.