

Ejemplo comparación de resultados predictores in silico

Cambio de estudio AR c.2450-118A>G (chr X:66942551 A/G, NM_000044: c.2450-118A>G)

Exón 7 e intrones adyacentes:

```
tgtgaatgacataaaaagccttaggtcctaaaaaatgagatctgagaagagtagaaaaaga  
aaagggttcaggaaatttgatttacttgactcctttcagatcggatcagctatcctttcc  
cctgagatctccctgacagactgaaagccccaagcacacagacttcaactaacagggaagc  
caatgtagatgggttccctgtggtgggtgggtcaagtctgtggcagaaaaacttgggtgctt  
tgtctaatagtctccttcgtgggcatgtctccctctccattctgtcttcatcccatcag  
TCCAGTGGATGGGCTGAAAAATCAAAATTCTTTGATGAACCTCGAATGAACACATCA  
AGGAACTCGATCGTATCATTCGATGCAAAAGAAAAAATCCACATCTGCTCAAGACGCT  
TCTACAGCTCACCAAGCTCCTGGACTCCGTGCAGCCT  
gtaagcaaacgatggagggtgcttatcagggagaacagcctgataagagcaatgataat  
atgcttctctagagtctggcaccacctgttgggagggtgcttccattccctctggctttg  
agtgtggtccaggaagaaaaatgtggtgaagaaaagaacaagggtcacagtgtcccagctg  
gatattgtgaaaggggtggaggagttagaacagagcagtttggaactcaggggaaggact  
tgcagcagatgaattctctaggcagacaaaacagacctggatgttttctcctctctctt
```

Se ha descrito que este cambio causa la producción de dos transcritos distintos: (1) inclusión exón críptico de 85 pb y (2) inclusión exón críptico de 202 pb. La mutación crea un sitio de unión nuevo para el *splicing enhancer*, proteína SRSF1 (SF2/ASF), lo que lleva a la inclusión de un pseudo-exón que contiene un codón de parada prematuro¹.

El cambio se encuentra antes del exón 7 (la **a** en color azul claro).

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

1. Käsäkoski J, Jääskeläinen J, Jääskeläinen T et al (2016) Complete androgen insensitivity syndrome caused by a deep intronic pseudoexon-activating mutation in the androgen receptor gene. Sci Rep 6:32819

NetGene2

Donor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
459	0	+	0.60	CGTGCAGCCT	^	GTAAGCAAAC	

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
433	326	1	-	0.42	CAGGAGCTTG	^	GTGAGCTGGT	

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
126	2	+	0.15	TTCCCCTGAG	^	ATCTCCCTGA	
300	1	+	0.43	CCCACATCAG	^	TTCCAGTGGA	
306	1	+	0.34	TCAGTTCCAG	^	TGGATGGGCT	
553	0	+	0.07	CTGTTGGGAG	^	GTGCTTCCAT	
580	2	+	0.18	TGGCTTTGAG	^	TGTGGTCCAG	
590	0	+	0.22	TGTGGTCCAG	^	GAAGAAAATG	
594	1	+	0.17	GTCCAGGAAG	^	AAAATGTGGT	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
457	302	1	-	0.77	TTGCTTACAG	^	GCTGCACGGA	
196	563	1	-	0.43	ACCCCCACAG	^	GGAACCATCT	
41	718	0	-	0.00	CTCTTCTCAG	^	ATACTCATTT	
26	733	-	-	0.00	CATTTTTTAG	^	GACCTAAGCT	

Donor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
184	2	+	0.55	AGGAAGCCAG	^	GTAGATGGTT	
459	0	+	0.60	CGTGCAGCCT	^	GTAAGCAAAC	

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
433	326	1	-	0.42	CAGGAGCTTG	^	GTGAGCTGGT	

Acceptor splice sites, direct strand

pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
126	2	+	0.15	TTCCCCTGAG	^	ATCTCCCTGA	
300	1	+	0.43	CCCACATCAG	^	TTCCAGTGGA	
306	1	+	0.34	TCAGTTCCAG	^	TGGATGGGCT	
553	0	+	0.07	CTGTTGGGAG	^	GTGCTTCCAT	
580	2	+	0.18	TGGCTTTGAG	^	TGTGGTCCAG	
590	0	+	0.22	TGTGGTCCAG	^	GAAGAAAATG	
594	1	+	0.17	GTCCAGGAAG	^	AAAATGTGGT	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
457	302	1	-	0.77	TTGCTTACAG	^	GCTGCACGGA	
196	563	1	-	0.49	ACCCCCACAG	^	GGAACCATCT	
41	718	0	-	0.00	CTCTTCTCAG	^	ATACTCATTT	
26	733	-	-	0.00	CATTTTTTAG	^	GACCTAAGCT	

Aparece un nuevo sitio *donor* en la secuencia mutante. Si empleara el sitio aceptor anterior (en verde), podría generarse la inclusión de un exón críptico de 57 pb.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 85.53.15.54.11443.0 :

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

Donor site predictions for 85.53.15.54.11454.0 :

Start	End	Score	Exon	Intron
177	191	0.53	aagccag	gtagatgg

Acceptor site predictions for 85.53.15.54.11443.0 :

Start	End	Score	Intron	Exon
78	118	0.98	gatttacttgactcctttc	agatcggatccagctatccttt
119	159	0.73	ccccctgagatctccctgac	agactgaaggccccaagcacac
280	320	0.47	tctgtcttcatccacatc	agttccagtggatgggctgaaa

Acceptor site predictions for 85.53.15.54.11454.0 :

Start	End	Score	Intron	Exon
78	118	0.98	gatttacttgactcctttc	agatcggatccagctatccttt
119	159	0.73	ccccctgagatctccctgac	agactgaaggccccaagcacac
280	320	0.47	tctgtcttcatccacatc	agttccagtggatgggctgaaa

Aparece un sitio *donor* en la secuencia mutante que no se predice en la secuencia wt. Si se empleara con el aceptor anterior (en rojo), se produciría la inclusión de un exón críptico de 46 pb. Si se empleara con el más alejado (en verde), se produciría la inclusión de un exón críptico de 87 pb.

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

MAXENT: -9.11 MDD: -5.71 MM: -6.08 WMM: -7.83 MAXENT: -19.14 MM: -20.27 WMM: -24.10

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
agcca(a/g)gtaga	gcccaag	gccagg	29904	73%

En el análisis del efecto del cambio, se obtiene una puntuación elevada (73%) para el cambio A>G, por lo que puede estar afectando al *splicing*.

CRYP-SKIP

Se emplea para el exón y las regiones flanqueantes a este, por lo que este predictor no va a ser útil para una variante intrónica profunda.

Human Splicing Finder

 New Acceptor splice site		Activation of a cryptic Acceptor site. Potential alteration of splicing	
Algorithm/Matix	position	sequences	variation
HSF Acceptor site (matrix AG)	chrX:67722698	- REF : ACAGGAAGCCAAGT - ALT : ACAGGAAGCCAAGT	42.53 > 70.4 => 65.53%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	19	466	actctaac	-1.79814887453	0.488069414317	69	10	18	-4.3445492
wt	19	461	aactcaacg	2.152245708	0.489035087719	64	10	18	-2.4809795
wt	19	420	gggtcgaag	-2.20553167138	0.501204819277	23	10	18	-1.5880937
wt	19	408	tggtcagaa	-2.40715792254	0.501240694789	11	10	18	-0.90744849
wt	19	386	cttctaagt	2.70518768944	0.501312335958	1	11	19	1.7365987
wt	19	344	tctttcatcc	-0.834687859501	0.469026548673	41	6	17	-2.2104219
wt	19	334	acatcagtt	-3.48383438876	0.462006079027	31	6	17	-2.6169722
wt	19	315	ggctgaaaa	-0.00644104793682	0.464516129032	12	6	17	-0.051929495
wt	19	308	aaatcaaaa	-3.3849348633	0.468646864686	5	6	17	-0.93034717
wt	19	296	ctttgatga	-0.0624952665958	0.46735395189	61	15	22	-3.1279994
wt	19	293	tgatgaact	-0.48114207014	0.46875 58	15	22	-3.1015735	
wt	19	282	gaatgaact	-0.127719277643	0.469314079422	47	15	22	-2.2667281
wt	19	273	acatcaagg	-3.76328350847	0.466417910448	38	15	22	-3.1214738
wt	19	255	gtatcattg	-2.41086135795	0.468 20	15	22	-1.4520558	
wt	19	220	tgtccaaga	-0.345384386203	0.46511627907	4	10	15	0.30330688
wt	19	201	agctcacca	1.65082763829	0.454081632653	46	9	15	-1.5771751
wt	19	171	ctttaagca	-0.148699627237	0.433734939759	16	9	15	-0.38939684
wt	19	149	gcctttatca	-3.73383985225	0.444444444444	36	9	19	-3.0183928
wt	19	146	ttatcaggcg	-3.408828125	0.439716312057	33	9	19	-2.7027675
wt	19	131	gccgtgagtg	1.42372724832	0.460317460317	18	9	19	0.14553471
wt	19	119	caatgataaa	-0.373571257971	0.473684210526	6	9	19	0.20570432
wt	19	116	tgtaaatat	-0.454292280883	0.477477477477	3	9	19	0.3652183
wt	19	54	ctttgatgtg	-0.745617230072	0.285714285714	49	0	0	-2.8995069
wt	19	27	tggtagaaga	-0.945833575188	0.272727272727	22	0	0	-1.2730408

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
mut	19	466	actctaac	-1.79814887453	0.488069414317	69	10	18	-4.3445492
mut	19	461	aactcaacg	2.152245708	0.489035087719	64	10	18	-2.4809795
mut	19	420	gggtcgaag	-2.20553167138	0.501204819277	23	10	18	-1.5880937
mut	19	408	tggtcagaa	-2.40715792254	0.501240694789	11	10	18	-0.90744849
mut	19	386	cttctaagt	2.70518768944	0.501312335958	1	11	19	1.7365987
mut	19	344	tctttcatcc	-0.834687859501	0.469026548673	41	6	17	-2.2104219
mut	19	334	acatcagtt	-3.48383438876	0.462006079027	31	6	17	-2.6169722
mut	19	315	ggctgaaaa	-0.00644104793682	0.464516129032	12	6	17	-0.051929495
mut	19	308	aaatcaaaa	-3.3849348633	0.468646864686	5	6	17	-0.93034717
mut	19	296	ctttgatga	-0.0624952665958	0.46735395189	61	15	22	-3.1279994
mut	19	293	tgatgaact	-0.48114207014	0.46875 58	15	22	-3.1015735	
mut	19	282	gaatgaact	-0.127719277643	0.469314079422	47	15	22	-2.2667281
mut	19	273	acatcaagg	-3.76328350847	0.466417910448	38	15	22	-3.1214738
mut	19	255	gtatcattg	-2.41086135795	0.468 20	15	22	-1.4520558	
mut	19	220	tgtccaaga	-0.345384386203	0.46511627907	4	10	15	0.30330688
mut	19	201	agctcacca	1.65082763829	0.454081632653	46	9	15	-1.5771751
mut	19	171	ctttaagca	-0.148699627237	0.433734939759	16	9	15	-0.38939684
mut	19	149	gcctttatca	-3.73383985225	0.444444444444	36	9	19	-3.0183928
mut	19	146	ttatcaggcg	-3.408828125	0.439716312057	33	9	19	-2.7027675
mut	19	131	gccgtgagtg	1.42372724832	0.460317460317	18	9	19	0.14553471
mut	19	119	caatgataaa	-0.373571257971	0.473684210526	6	9	19	0.20570432
mut	19	116	tgtaaatat	-0.454292280883	0.477477477477	3	9	19	0.3652183
mut	19	54	ctttgatgtg	-0.745617230072	0.285714285714	49	0	0	-2.8995069
mut	19	27	tggtagaaga	-0.945833575188	0.272727272727	22	0	0	-1.2730408

No hay diferencias entre los resultados.

IntSplice

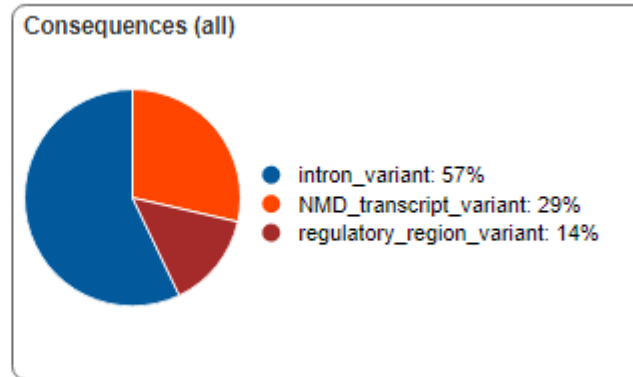
SNV at chrX:66942551 can't be predicted by IntSplice.

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	-
Overlapped genes	1
Overlapped transcripts	4
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. Al ser variante NMD, tiene que tener algún fallo en el mRNA para que se active este mecanismo degradador del mensajero.

Uploaded variant	Location	Allele	Consequence	Symbol	Gene	Feature type
ENST00000374690.3:c.2450-118A>G	X:67722709-67722709	G	intron_variant	AR	ENSG00000169083	Transcript
ENST00000374690.3:c.2450-118A>G	X:67722709-67722709	G	intron_variant NMD_transcript_variant	AR	ENSG00000169083	Transcript
ENST00000374690.3:c.2450-118A>G	X:67722709-67722709	G	intron_variant	AR	ENSG00000169083	Transcript
ENST00000374690.3:c.2450-118A>G	X:67722709-67722709	G	intron_variant NMD_transcript_variant	AR	ENSG00000169083	Transcript
ENST00000374690.3:c.2450-118A>G	X:67722709-67722709	G	regulatory_region_variant	-	-	RegulatoryFeature

ESEfinder

Cuando se buscan los posibles sitios de *splicing*, solo se obtienen dos resultados con la posición de interés con puntuación positiva en las matrices de 5'SS: ctaacaggaagccaagtagatggttcctg (5.54030 y 5.32370) y agtagatggttcctgtgggggtgggggtc (3.09710 y 3.41230). Si buscamos los resultados equivalentes en las predicciones para la secuencia mutada, vemos que, para la primera las puntuaciones suben bastante (8.92770 y 8.61260), mientras que para la segunda suben muy poco (3.33770 y 3.64510). Por otro lado, hay un resultado con puntuación positiva en las matrices 3'SS: gacttcaactaacaggaagccaagtagatg

(4.69740 y 4.34330). Su equivalente en la secuencia mutada tiene puntuaciones prácticamente iguales (4.68960 y 4.26320), por lo que no es relevante. Por lo tanto, se está generando un sitio 5'SS mucho más fuerte que el original (el primero), lo que hará que sea más probable la alteración del *splicing*.

EX-SKIP

Predictor para secuencias exónicas.

HOT-SKIP

Predictor para secuencias exónicas.