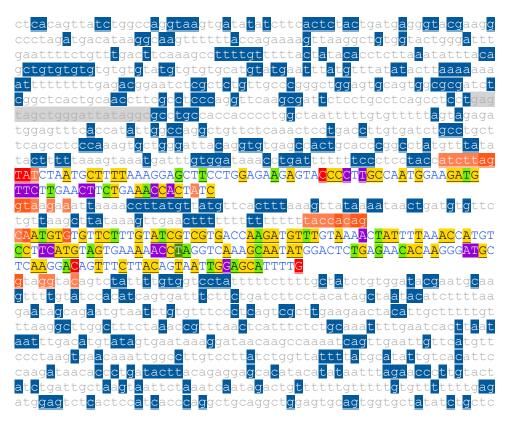
Ejemplo comparación de resultados predictores in sillico

Cambio de estudio: ATM c.2921+1G>A (chr11:108141874 G/A, rs587781558 o NM_000051.3:c.2921+1G>A)

Exones 19 y 20 e intrones adyacentes:



Se ha descrito que este cambio causa la pérdida completa del exón 21. Como consecuencia de la eliminación de 83 pb, el marco de lectura normal cambia después de la metionina 946, lo que produce un codón de parada prematuro en el residuo 955¹. La secuencia de ATM que ellos emplearon, al igual que la actualmente vigente, tiene un tamaño de 9168 pb² y el exón 21 con la nomenclatura actual tiene un tamaño de 76 pb, mientras que el exón 19 tiene 83 pb. Por lo tanto, el efecto será la pérdida del exón 19.

El cambio se encuentra en primera posición después del exón 19 (la g en color rojo).

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

- 1. Gilad, S.; Khosravi, R.; Shkedy, D.; Uziel, T.; Ziv, Y.; Savitsky, K.; ... Bar-Shira, A. (1996). Predominance of null mutations in ataxia-telangiectasia. Human Molecular Genetics, 5(4), 433–439.
- 2. Savitsky, K.; Sfez, S.; Tagle, D. A.; Ziv, Y.; Sartiel, A.; Collins, F. S.; ... Rotman, G. (1995). The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. Human Molecular Genetics, 4(11), 2025–2032.

NetGene2

The sequence: Wt has the	following composit:	tion:	The sequence: mut has the following composition:							
Length: 775 nucleotides. 26.1% A, 19.2% C, 18.3% G	, 36.4% T, 0.0% X,	, 37.5% G+C	Length: 775 nucleotides. 26.2% A, 19.2% C, 18.2% G, 36.4% T, 0.0% X, 37.4% G+C							
26.1% A, 19.2% C, 18.3% G, 36.4% T, 0.0% X, 37.5% G+C Donor splice sites, direct strand pos 5'->3' phase strand confidence 5' exon int 567 2 + 0.56 GAGCATTTTG^GTA Donor splice sites, complement strand pos 3'->5' pos 5'->3' phase strand confidence 5' exon int 215 561 1 - 0.36 TACTAAGATG^GTA Acceptor splice sites, direct strand pos 5'->3' phase strand confidence 5' intron exc 223 0 + 0.36 ACCATCTTAG^TAT 241 0 + 0.27 GCTTTTAAAGGAG^CTT 244 0 + 0.07 TTTAAAGGAG^CCTT 410 2 + 0.65 TTTACCACAG^CCAA			Donor splice sites, direct strand							
	•		pos 5'->3' phase strand confidence 5' exon intron 3' 567 2 + 0.56 GAGCATTTTG^GTAGGTACAG							
Donor splice sites, compl	ement strand		Donor splice sites, complement strand							
the state of the s	•		pos 3'->5' pos 5'->3' phase strand confidence 5' exon intron 3' 215 561 1 - 0.36 TACTAAGATG^GTAGGAGGGA							
			Acceptor splice sites, direct strand							
pos 5'->3' 223 241 244 410	phase strand cor 0 + 0 + 0 + 2 +	0.36 ACCATCTTAG^TATCTAATGC 0.27 GCTTTTAAAG^GAGCTTCCTG 0.07 TTTAAAGGAG^CTTCCTGGAG 0.65 TTTACCACAG^CAATGTGTGT	Acceptor splice sites, direct strand pos 5'->3' phase strand confidence 5' intron exon 3' 223 0 + 0.32 ACCATCTTAG^TATCTAATGC 241 0 + 0.23 GCTTTTAAAG^GAGCTTCCTG 410 2 + 0.65 TTTACCACAG^CAATGTGTGT 751 - + 0.00 TTTGTTTAAG^GCTTGGCTTT							
pos 5'->3' 223 241 244 410 751	phase strand com 0 + 0 + 0 + 2 + - +	0.36 ACCATCTTAG^TATCTAATGC 0.27 GCTTTTAAAG^GAGCTTCCTG 0.07 TTTAAAGGAG^CTTCCTGGAG 0.65 TTTACCACAG^CAATGTGTGT	pos 5'->3' phase strand confidence 5' intron exon 3' 223 0 + 0.32 ACCATCTTAG^TATCTAATGC 241 0 + 0.23 GCTTTTAAAG^GAGCTTCCTG 410 2 + 0.65 TTTACCACAG^CAATGTGTGT							

Al comparar los resultados que se obtienen del análisis de la secuencia *wild type* (wt) y la secuencia con la mutación (mut), se observa que se pierde un sitio *aceptor* en la hebra *direct* de la secuencia mutante. El sitio que desaparece está dentro del exón, que no es empleado en el *splicing* normal (entre chr11:108141802 y chr11: 108141811) y, por lo que no tendrá efecto en el *splicing*.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 80.102.134.106.14034.0:

Sta	rt	End	Score	Exon Intron
	13	27	1.00	tggccag gt aagtga
	45	59	0.79	gatgagg gt acgaag
3	21	335	0.40	ctcccag gt tcaagc
5	02	516	0.85	attacag gt gtgagc
6	77	691	0.85	cactatc gt aagaaa

0.98

951

937

Donor site predictions for 80.102.134.106.14049.0:

C44	End	C	From Todayan
Start	Ena	Score	Exon Intron
13	27	1.00	tggccag gt aagtga
45	59	0.79	gatgagg gt acgaag
321	335	0.40	ctcccag gt tcaagc
502	516	0.85	attacag gt gtgagc
937	951	0.98	cattttg gt aggtac

Acceptor site predictions for 80.102.134.106.14034.0:

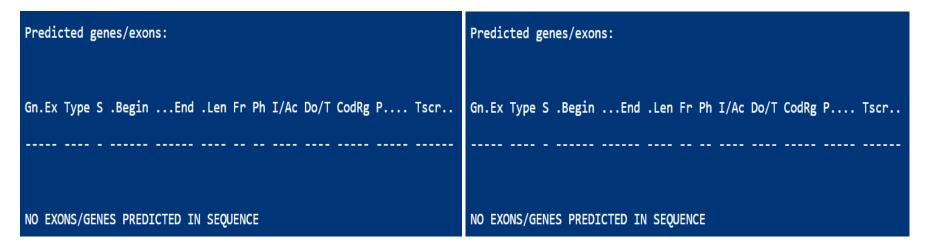
cattttg**gt**aggtac

Acceptor site predictions for 80.102.134.106.14049.0:

Start	End	Score	Intron	Exon	Start	End	Score	Intron	Exon
307	347	0.89		gcctccc ag gttcaagcgattctcctgcc	307	347	0.89	actgcaaccttcgc	ctccc ag gttcaagcgattctcctgcc
331	371	0.42	_	ctgcctc ag cctcctgagtagctgggatt	224	371	0.42	caagcgattctcct	gcctc ag cctcctgagtagctgggatt
394	434	0.88		tgttttt ag tagagatggagtttcaccat		434	0.88	ggctaattttgtg	ttttt ag tagagatggagtttcaccat
580	620	0.42		ccatctt ag tatctaatgcttttaaagga		620	0.42	ttttccctcctacc	atctt ag tatctaatgcttttaaagga
767	807	0.97		ttaccac ag caatgtgtgttctttgtatc		807	0.97		accac ag caatgtgtgttctttgtatc
1027	1067	0.73		cctacat ag ctaatacatcttttaagaat		1067	0.73		tacat ag ctaatacatcttttaagaat
	1115	0.79	_	ttccctc ag tcgcttgaagaactacattg			0.79		ccctc ag tcgcttgaagaactacattg
	1148	0.95		ttgttta ag gcttggctttctaaaccgtt			0.95		gttta ag gcttggctttctaaaccgtt
		0.45		_					
1427	1467	0.45	gagteteactee	atcaccc ag gctgcaggctggagtgcagt	1427	1467	0.45	gagtctcactccat	caccc ag gctgcaggctggagtgcagt

Al comparar los resultados que se obtienen del análisis de la secuencia *wild type* (izquierda) y la secuencia con la mutación (derecha), se observa que se pierde un sitio *donor* en la secuencia mutante. Este es el sitio *donor* del exón 19 (chr11:108141867-108141881), por lo que en vez de tener un exón 19 y otro 20, el *spliceosome* reconocerá el sitio *aceptor* del exón 19 y el *donor* del exón 20, generándose un exón más largo de 343 bp (sumando las pb de E19, I19 y E20). También podría perderse el exón completo.

GENSCAN → no da resultados para este cambio



MaxEntScan (solo se estudia 5' por la proximidad con el cambio)

>wt

AACTTCTGAAACCACTATCgtaagaaattaaaac MAXENT: -11.78 MDD: -9.83 MM: -9.88 WMM: -7.46

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

Spliceman

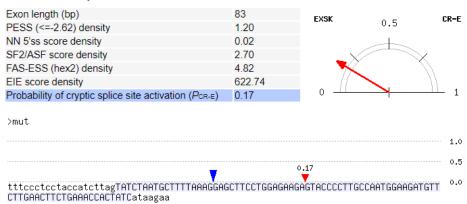
TGAACTTCTGAAACCACTATC(g/a)taagaaattaaaacct

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ctatc(g/a)taaga	togtaa	tcataa	29076	68%

En el análisis de la región adyacente al cambio, se obtiene una puntuación elevada (68%) para el cambio G>A, por lo que puede estar afectando al splicing.

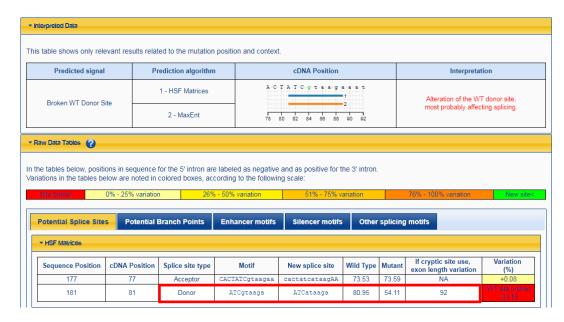
CRYP-SKIP

Results for sequence mut



Parece que hay un sitio críptico de *splicing* dentro del propio exón, pero el cambio de interés (la primera **a** en minúsculas después 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 (version Antigua)



Se podría producir una alteración en el sitio *donor* de la secuencia *wild type*, que afectaría al *splicing*. Si se empleara el nuevo sitio críptico (ATCataaga), habría una variación de 92 pb en la longitud del exón. El sitio nuevo de *donor* se empleará en un 54.11% de las veces, mientras que el *wild type* lo se utilizará en el 80.95% de las veces que se lleve a cabo el *splicing*. Por lo tanto, se podría estar perdiendo el exón 19, aunque este tenga una longitud de 83, por lo que se perdería algo más (parte del intrón) a parte del exón.

SVM-BPfinder

									Ī	i									
seq_id		ss_dist bp_s		y_cont ppt_off						seq_id	agez	ss_dist	bp_seq bp_scr	y_cont ppt_off	ppt_len ppt_scr	svm_scr			
wt	28			-3.07747864696				18	-0.6448103	mut	28	250	tgtttatat	-3.07747864696	0.559183673469	3	8	18	-0.6448103
wt	28			-4.30072364603				51	-2.084864	mut	28	239	tttttaaag	-4.30072364603	0.551282051282	23	27	51	-2.084864
wt	28		0	-1.42501787682		22		51	-0.89482685	mut	28	238	ttttaaagt	-1.42501787682	0.55364806867	22	27	51	-0.89482685
wt	28	_	0	-0.55251643678				51	-0.23420493	mut	28	233	aagtaaatg	-0.55251643678	0.561403508772	17	27	51	-0.23420493
wt	28		0	-1.05449072678				51	-0.17576232	mut	28	229	aaatgattt	-1.05449072678	0.566964285714	13	27	51	-0.17576232
wt	28	-		-0.811562272927					0.61751222	mut	28	218	ggataaacc	-0.811562272927	0.572769953052	2	27	51	0.61751222
wt	28		-0	0.245597469235		_			1.0674723	mut	28	212	acctgattt	0.245597469235	0.574879227053	1	22	48	1.0674723
wt	28			-2.54600400235				15	-0.85127799	mut	28	189	atcttagta	-2.54600400235	0.538043478261	9	7	15	-0.85127799
wt	28				0.539772727273	_			1.5473316	mut	28	181	atctaatgc	2.28525063653	0.539772727273	1	7	15	1.5473316
wt	28			-4.12603823658				17	-2.2768572	mut	28	173	cttttaaag	-4.12603823658	0.529761904762	22	11	17	-2.2768572
wt	28		00	-1.53245739368				17	-1.1970249	mut	28	172	ttttaaagg	-1.53245739368	0.532934131737	21	11	17	-1.1970249
wt	28		•	-0.428095221544		43		18	-2.1366026	mut	28	123	tcttgaact	-0.428095221544	0.567796610169	43	10	18	-2.1366026
wt	28		-0		0.563636363636	35		18	-1.0131341	mut	28	115	ttctgaaac	1.15134512884	0.563636363636	35	10	18	-1.0131341
wt	28			-1.68650853105		21		18	-1.2351142	mut	28	104	ctatcataa	-2.87305638291	0.565656565657	24	10	18	-1.8919435
wt	28			-3.55055560732				18	-1.5145929	mut	28	101	tcataagaa	-2.48718769318	0.572916666667	21	10	18	-1.5486179
wt	28			-0.709023521857				18	-0.33651488	mut	28	94	aaattaaaa	-3.55055560732	0.595505617978	14	10	18	-1.5145929
wt	28			-1.15931800217				18	-0.068877299	mut	28	93	aattaaaac	-0.709023521857	0.602272727273	13	10	18	-0.33651488
wt	28	_		-2.50541896492				18	-0.27934294	mut	28	86	accttatgt	-1.15931800217	0.604938271605	6	10	18	-0.068877299
wt	28			0.458389917938		11		-0.58574		mut	28	81	atgttatgt	-2.50541896492	0.605263157895	1	10	18	-0.27934294
wt	28			-4.90201693136				17	-2.3730786	mut	28	75	tgttcactt	0.458389917938	0.6 24	11	17	-0.5857	4874
wt	28			-1.25033246736		11		-0.87702		mut	28	70	actttaaag	-4.90201693136	0.584615384615	19	11	17	-2.3730786
wt	28			-3.96710231184				17	-1.6189721	mut	28	69	ctttaaagt	-1.25033246736	0.59375 18	11	17	-0.8770	2296
wt	28				0.625 10	11		-0.51397		mut	28	64	aagttataa	-3.96710231184	0.610169491525	13	11	17	-1.6189721
wt	28				0.66666666667	5			0.92716836	mut	28	61	ttataaaat	-1.64220173827	0.625 10	11	17	-0.5139	7891
wt	28		-00-		0.68085106383	1			1.4676619	mut	28	56	aaataactg	1.19576613069	0.66666666667	5	11	17	0.92716836
wt	28				0.685714285714			49	-1.1174366	mut	28	52	aactgatgt	1.91782225232	0.68085106383	1	11	17	1.4676619
wt	28		0		0.705882352941				0.1126757	mut	28	40	ctgttaagc	-2.8632176691	0.685714285714	17	19	49	-1.1174366
wt	28			-2.38394267454				49	-0.53757816	mut	28	39	tgttaagct	0.100154611789	0.705882352941	16	19	49	0.1126757
wt	28		•	-1.58005402731				49	-0.018359155	mut	28	34	agcttataa	-2.38394267454	0.724137931034	11	19	49	-0.53757816
wt	28	25 agtt	tgaact	-0.445827552767	0.9 2	19	49	0.847768	867	mut	28	31	ttataaagt	-1.58005402731	0.769230769231	8	19	49	-0.018359155
										mut	28	25	agttgaact	-0.445827552767	0.9 2	19	49	0.84776	867
										I									

Al comparar ambas secuencias (*wild type* a la izquierda, mutante a la derecha), desaparece el *branch point* de WT (chr11:108141872-108141880) y aparecen dos nuevos en la mutante (chr11:108141869-108141877 y chr11:108141872-108141880). Tienen puntuaciones negativas, que según el programa generalmente no son consideradas. Se estará utilizando un nuevo sitio de *splicing* (el mismo que predice HSF).

IntSplice

SNV at chr11:108141874 can't be predicted by IntSplice.

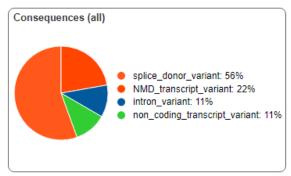
Prediction shows either Abnormal or Normal.

Prediction Genomic Mutation Ensembl 64 Transcript ID and Exon No. No Se obtienen resultados para esta variante.

Input queries: hg19, chr11, 108141874

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	6
Overlapped regulatory features	0



El predictor encuentra 17 resultados para la variante (rs587781558). Solo nos interesan los resultados que indiquen el cambio G>A (6). En cinco de estos, nos indican que es una variante que está afectando al sitio donor del splicing y dos de ellos que está implicada en el NMD o non-sense mediated decay (fenómeno que consiste en evitar la producción de proteínas truncadas con efectos deletéreos³).

Uploaded variant	Location	Allele	Consequence	Symbol	<u>Gene</u>	Feature type	Feature	Biotype	Existing variant
rs587781558	<u>11:108271147-</u> <u>108271147</u>	A	splice donor variant	ATM	ENSG00000149311	Transcript	ENST00000278616.8	protein_coding	rs587781558, CS991304, COSV53740769, COSV53748775, COSV53767471
rs587781558	<u>11:108271147-</u> <u>108271147</u>	A	intron_variant, non_coding_transcript_variant	ATM	ENSG00000149311	Transcript	ENST00000419286.2	processed_transcript	rs587781558, CS991304, COSV53740769, COSV53748775, COSV53767471
rs587781558	<u>11:108271147-</u> <u>108271147</u>	A	splice donor variant	ATM	ENSG00000149311	Transcript	ENST00000452508.6	protein_coding	rs587781558, CS991304, COSV53740769, COSV53748775, COSV53767471
rs587781558	<u>11:108271147-</u> <u>108271147</u>	A	splice donor variant, NMD_transcript_variant	ATM	ENSG00000149311	Transcript	ENST00000527805.6	nonsense_mediated_decay	rs587781558, CS991304, COSV53740769, COSV53748775, COSV53767471
rs587781558	<u>11:108271147-</u> <u>108271147</u>	A	splice donor variant, NMD_transcript_variant	ATM	ENSG00000149311	Transcript	ENST00000675595.1	nonsense_mediated_decay	rs587781558, CS991304, COSV53740769, COSV53748775, COSV53767471
rs587781558	<u>11:108271147-</u> <u>108271147</u>	A	splice_donor_variant	ATM	ENSG00000149311	Transcript	ENST00000675843.1	protein_coding	rs587781558, CS991304, COSV53740769, COSV53748775, COSV53767471

Nota: las coordenadas no coinciden porque el buscador emplea la versión 38 del genoma y las coordenadas al principio del fichero son las de la versión 37.

^{3.} Hug, N.; Longman, D.; & Cáceres, J. F. (2015). Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Research, 44(4), 1483–1495.

ESEfinder



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 5'SS en un único resultado (el mostrado en la imagen) por lo que esta deberá ser la región que produzca este sitio de *splicing*. Si se hace la búsqueda para la secuencia mutada, no aparece ningún resultado positivo con la posición de interés. Por lo tanto, se está perdiendo el sitio *donor*, afectando el *splicing*.

EX-SKIP

EX-SKIP - Results for submitted sequences

Seq	PESS (count)	FAS-ESS hex2 (count)	hex3	IIE (count)	IIE	NI-ESS trusted	NI-ESS all	PESE (sount)	RESCUE -ESE	EIE (count)	EIE (oum)	NI-ESE trusted	NI-ESE all	ESS (total)	ESE (total)	ESS/ESE
	(count)	(count)	(count)	(count)	(sum)	(count)	(sum)	(count)	(count)	(count)	(sum)	(count)	(sum)	(total)	(total)	(ratio)
wt	1	3	2	19	263.8061	12	-15.6804	9	19	29	516.8703	33	39.1610	37	90	0.41
mut	1	3	2	19	263.8061	12	-15.6804	9	19	29	516.8703	33	39.1610	37	90	0.41

Both alleles have a comparable chance of exon skipping.

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

HOT-SKIP

HOT-SKIP - Results for submitted sequences

Note: only the first sequence was analyzed!

>ex19
tcttagTATCTAATGCTTTTAAAGGAGCTTCCTGGAGAAGAGTACCCCTTGCCAATGGAAGATGTTCTTGAACTTCTGAA
ACCACTATCgtaa

Mutation(s) E+10C>T, E+10C>G and E+2A>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.