

Ejemplo comparación de resultados predictores in silico

Cambio de estudio DIDO1 c.2214+2T>C (chr20:62896231 T/C, o NM_001193369.2: c.2214+2T>C)

Exón 8 e intrones adyacentes:

gtgggctcttctgcttgatttcctaccatataaagtgaaggccttcattagagtgtag
ctcgaggattgctcactgttcaaaaagtttaatatatacaaaagtgtcttatgtagtctct
tttttttattctgttag
AGTCAATGACAGCGATGACTTAATCATGACAGAAAACGAAGTAGGAAAATTCCCCTC
TATTGACAGAGGAGATGTTTAACTTGTTCAAGTTACAGATAATCGTACAAGAGTAAATA
TCGCAGCATCATGTTC AACCTGAAGGACCTTAAAAATCAG
gtgtgtgctcgtgtgtcgtgcatgtgagattctgttctgttcgtcgtgggtgctaaagg
atcaaccaatatatttgagatgttcgtgtcctttccaggaaagctaaagcgtttctttctc
ctctgtgatcctctctgcgcacagcatttgttgctcagtcgttgacgtctctggaaaccc
tcttcccttcctgtctaaactagggctgcgtcgtctcctggctctcctaagtcacct
tttatcaccacattagttctctgttaacattgctgtaagggtacaagccagtcgcttcac
tgggtccaaaaataaagtcagttctcatccagtgtaggggtgtcgtattccggc
ctcagacagcctttttccaccttattttctcatattctcatcagccgagtgagcctgatg

El cambio se encuentra en la segunda posición justo después del exón 8 (la **t** 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'
299	0	+		0.94		TAAAAATCAG	GTGTGTGCC	H
579	0	+		0.31		TGCTGTAAAG	GTACAAGCCA	

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
26	693	-	-	0.00		ATTTTAATGG	GTAAGAAATC	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
103	0	+		0.07		TATACACAAG	TGTCTTATGT	
115	0	+		0.19		TCTTATGTAG	TTCTTTTTTT	
138	2	+		1.00		TTCTGTTT	CAGAGTCAATGAC	H
140	1	+		0.44		CTGTTT	CAGAGTCAATGAC	
150	2	+		0.31		TCAATGACAG	CGATGACTTA	
236	1	+		0.26		CAAGTTACAG	ATAATCGCTA	
501	0	+		0.15		CCTTAAC	TAGGGTCGCGTCC	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
278	441	0	-	0.69		GGTCCTTCAG	GTTGAACATG	
244	475	1	-	0.16		ACTCTTGTAG	CGATTATCTG	
194	525	0	-	0.18		CAATATGGAG	GGCAATTTTT	

Donor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	exon	intron	3'
579	0	+		0.31		TGCTGTAAAG	GTACAAGCCA	

Donor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	exon	intron	3'
26	693	-	-	0.00		ATTTTAATGG	GTAAGAAATC	

Acceptor splice sites, direct strand

pos	5'→3'	phase	strand	confidence	5'	intron	exon	3'
103	0	+		0.07		TATACACAAG	TGTCTTATGT	
115	0	+		0.19		TCTTATGTAG	TTCTTTTTTT	
138	2	+		1.00		TTCTGTTT	CAGAGTCAATGAC	H
140	1	+		0.44		CTGTTT	CAGAGTCAATGAC	
150	2	+		0.25		TCAATGACAG	CGATGACTTA	
236	1	+		0.26		CAAGTTACAG	ATAATCGCTA	
501	0	+		0.15		CCTTAAC	TAGGGTCGCGTCC	

Acceptor splice sites, complement strand

pos 3'→5'	pos 5'→3'	phase	strand	confidence	5'	intron	exon	3'
278	441	0	-	0.57		GGTCCTTCAG	GTTGAACATG	
194	525	0	-	0.18		CAATATGGAG	GGCAATTTTT	

Se pierde uno de los sitios de *donor* predichos en la secuencia *wild type*. Este coincide con el sitio *donor* del exón, por lo que es muy probable que esté afectando al *splicing*, perdiéndose el exón completo o, si se empleara el otro *donor*, se incluiría un exón críptico.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.3.123.423500.0 :

Start	End	Score	Exon	Intron
245	259	0.52	tacaaga	gtaaatat
292	306	0.98	aaatcag	gtgtgtgc
572	586	0.89	tgtaaag	gtacaagc

Donor site predictions for 10.42.0.139.423674.0 :

Start	End	Score	Exon	Intron
245	259	0.52	tacaaga	gtaaatat
572	586	0.89	tgtaaag	gtacaagc

Acceptor site predictions for 10.42.3.123.423500.0 :

Start	End	Score	Intron	Exon
118	158	0.98	ccttttttttattctgttc	agagtcaatgacagcgatgact
216	256	0.53	ttaacttgtttcaagttac	agataatcgctacaagagtaaa
375	415	0.91	agatgttcgtgtcctttcc	aggaaagctaagacgtttcttt
481	521	0.69	tcccttcctgccttaact	agggtcgcgtccgcttcctggc

Acceptor site predictions for 10.42.0.139.423674.0 :


Start	End	Score	Intron	Exon
118	158	0.98	ccttttttttattctgttc	agagtcaatgacagcgatgact
216	256	0.53	ttaacttgtttcaagttac	agataatcgctacaagagtaaa
375	415	0.91	agatgttcgtgtcctttcc	aggaaagctaagacgtttcttt
481	521	0.69	tcccttcctgccttaact	agggtcgcgtccgcttcctggc

Se pierde uno de los sitios de *donor* predichos en la secuencia *wild type*. Este coincide con el sitio *donor* del exón, por lo que es muy probable que esté afectando al *splicing*, perdiéndose el exón completo, perdiéndose parte del exón (si se usa el *donor* anterior), disminuyendo el tamaño del exón de 160 a 113 pb; o aumentando el tamaño del exón (si se usa el *donor* posterior), aumentando el tamaño del exón de 160 a 443.

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
tcagg(t/c)gtgtg	tgtgtg	cgtgtg	31552	84%

Human Splicing Finder

Type	↑↓	Interpretation	↑↓
 Broken WT Donor Site		Alteration of the WT Donor site, most probably affecting splicing	

Algorithm/Matix	position	sequences	variation
HSF Donor site (matrix GT)	chr20:62896235	- REF : CAGG T GTGT - ALT : CAGG C GTGT	89.96 > 62.82 => -30.17%
MaxEnt Donor site	chr20:62896235	- REF : CAGG T GTGT - ALT : CAGG C GTGT	6.99 > -0.76 => -110.87%

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	19	63	gagtaaata		-0.963398652921	0.48275862069	45	9	10 -2.5747873
wt	19	49	gcatcatgt		-1.39668286914	0.477272727273	31	9	10 -1.8600339
wt	19	43	tgttcaacc		-0.188948690015	0.447368421053	25	9	10 -1.0170172
wt	19	37	acctgaagg		-0.185817476477	0.4375	19	9	10 -0.63918843
wt	19	27	ccctaaaaa		0.54832918423	0.454545454545	9	9	10 0.28675309
wt	19	21	aaatcaggt		-3.04574808516	0.5	3	9	10 -0.72602759
mut	19	63	gagtaaata		-0.963398652921	0.48275862069	58	0	0 -3.490824
mut	19	49	gcatcatgt		-1.39668286914	0.477272727273	44	0	0 -2.7760705
mut	19	43	tgttcaacc		-0.188948690015	0.447368421053	38	0	0 -1.9330538
mut	19	37	acctgaagg		-0.185817476477	0.4375	32	0	0 -1.555225
mut	19	27	ccctaaaaa		0.54832918423	0.454545454545	22	0	0 -0.62928354
mut	19	21	aaatcaggc		-2.70511060235	0.5	16	0	0 -1.5086886

La única diferencia que hay entre ambas es que en el último BP que detecta cambia el último nucleótido a consecuencia de la mutación que se produce, pero, como ambas predicciones tienen puntuación negativa, no se van a tener en cuenta.

Variant Effect Predictor tool

ENST00000395340.5:c.2214+2T>C	20:62896231-62896231 G	splice_donor_variant	DIDO1	ENSG00000101191 Transcript	ENST00000266070.8 protein_coding	-	-	-	-	-	-	-
ENST00000395340.5:c.2214+2T>C	20:62896231-62896231 G	splice_donor_variant	DIDO1	ENSG00000101191 Transcript	ENST00000395340.5 protein_coding	-	-	-	-	-	-	-
ENST00000395340.5:c.2214+2T>C	20:62896231-62896231 G	splice_donor_variant	DIDO1	ENSG00000101191 Transcript	ENST00000395343.6 protein_coding	-	-	-	-	-	-	-

ESEfinder

Se observan tres predicciones que tienen puntuaciones positivas para las matrices 5':

284 (-435)	GACCCATAAAATCAGgtgtgtgcctgggtgt	11.53130	284 (-435)	GACCCATAAAATCAGgtgtgtgcctgggtgt	-0.67080	284 (-435)	GACCCATAAAATCAGgtgtgtgcctgggtgt	11.50690	284 (-435)	GACCCATAAAATCAGgtgtgtgcctgggtgt	-1.23450
286 (-433)	CCCTATAAAATCAGgtgtgtgcctgggtgtgt	2.17680	286 (-433)	CCCTATAAAATCAGgtgtgtgcctgggtgtgt	-32.80760	286 (-433)	CCCTATAAAATCAGgtgtgtgcctgggtgtgt	2.23350	286 (-433)	CCCTATAAAATCAGgtgtgtgcctgggtgtgt	-32.73750
295 (-424)	TCAGgtgtgtgcctgggtgtcggtgcattgg	2.83240	295 (-424)	TCAGgtgtgtgcctgggtgtcggtgcattgg	-7.87220	295 (-424)	TCAGgtgtgtgcctgggtgtcggtgcattgg	2.80590	295 (-424)	TCAGgtgtgtgcctgggtgtcggtgcattgg	-9.07290

Cuando se comparan estas puntuaciones con las equivalentes en la secuencia mutante si observa que en el caso de 284 las puntuaciones descienden considerablemente, en 286 aumentan un punto y en 295 disminuyen más de medio:

284 (-435)	GACCCATAAAATCAGgctgtgtgcctgggtgt	4.59520	284 (-435)	GACCCATAAAATCAGgctgtgtgcctgggtgt	-1.13420	284 (-435)	GACCCATAAAATCAGgctgtgtgcctgggtgt	4.39680	284 (-435)	GACCCATAAAATCAGgctgtgtgcctgggtgt	-1.70000
286 (-433)	CCCTATAAAATCAGgctgtgtgcctgggtgtgt	1.27760	286 (-433)	CCCTATAAAATCAGgctgtgtgcctgggtgtgt	-31.00540	286 (-433)	CCCTATAAAATCAGgctgtgtgcctgggtgtgt	1.33460	286 (-433)	CCCTATAAAATCAGgctgtgtgcctgggtgtgt	-29.97250
295 (-424)	TCAGgctgtgtgcctgggtgtcggtgcattgg	3.49780	295 (-424)	TCAGgctgtgtgcctgggtgtcggtgcattgg	-8.30160	295 (-424)	TCAGgctgtgtgcctgggtgtcggtgcattgg	3.49700	295 (-424)	TCAGgctgtgtgcctgggtgtcggtgcattgg	-9.48310

Por otro lado se observa qué la puntuación para la predicción 288 pasa de ser negativa a positiva con una diferencia de más de 2 puntos:

288 (-431)	CTAAAAATCAGgtgtgtgtgcctgggtgtcggt	-2.02930	288 (-431)	CTAAAAATCAGgtgtgtgtgcctgggtgtcggt	-30.99410	288 (-431)	CTAAAAATCAGgtgtgtgtgcctgggtgtcggt	-1.93620	288 (-431)	CTAAAAATCAGgtgtgtgtgcctgggtgtcggt	-30.78880
288 (-431)	CTAAAAATCAGgctgtgtgtgcctgggtgtcggt	0.12150	288 (-431)	CTAAAAATCAGgctgtgtgtgcctgggtgtcggt	-29.40580	288 (-431)	CTAAAAATCAGgctgtgtgtgcctgggtgtcggt	0.16570	288 (-431)	CTAAAAATCAGgctgtgtgtgcctgggtgtcggt	-29.17510

En conclusión, lo más probable es que se está debilitando un sitio *donor* (284), afectando al *splicing*, porque es el que mayor diferencia de puntuaciones tiene.