Ejemplo comparación de resultados predictores in sillico

Cambio de estudio CLDN12 c.*581T>C (chr7:90413992 T/C, NM_001185072.3: c.*581T>C)

Exón 3 e intrones adyacentes:

El cambio se encuentra en la antepenúltima fila del 3'UTR después del exón 3 (la **t** en color gris).

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                                    Donor splice sites, direct strand
---------------------------------                                    ---------------------------------
          pos 5'->3'  phase strand  confidence 5'    exon intron   3'           pos 5'->3'  phase strand  confidence 5'    exon intron   3'
             273        2      +        0.37      AATGTGCCCG^GTATGACGGG              273        2      +        0.37      AATGTGCCCG^GTATGACGGG
             524        1      +        0.69      TTCCTGGCAG^GTACTGTGAG              524        1      +        0.69      TTCCTGGCAG^GTACTGTGAG
            1258        0      +        0.54      TTTATTTCAG^GTAAGGGTAA             1258        0      +        0.56      TTTATTTCAG^GTAAGGGTAA
            1536        -      +        0.00      TTTATGACAG^GTTAGTGTTT             1536        -      +        0.00      TTTATGACAG^GTTAGTGTTT

Donor splice sites, complement strand                                Donor splice sites, complement strand
-------------------------------------                                -------------------------------------
  pos 3'->5'  pos 5'->3'  phase strand  confidence 5'    exon intron   3'   pos 3'->5'  pos 5'->3'  phase strand  confidence 5'    exon intron   3'
    753         808         1      -        0.86      AGGGCTGTGA^GTAAGTATGC     753         808         1      -        0.86      AGGGCTGTGA^GTAAGTATGC

Acceptor splice sites, direct strand                                 Acceptor splice sites, direct strand
------------------------------------                                 ------------------------------------
          pos 5'->3'  phase strand  confidence 5'    intron exon   3'           pos 5'->3'  phase strand  confidence 5'    intron exon   3'
             172        1      +        0.19      CTCTTTGCAG^GGACTCTGCT              172        1      +        0.19      CTCTTTGCAG^GGACTCTGCT
             194        2      +        0.32      CCAACTGGAG^AAAATTACGA              194        2      +        0.32      CCAACTGGAG^AAAATTACGA
             221        2      +        0.19      CATTCAACAG^AAACGAGAAG              221        2      +        0.19      CATTCAACAG^AAACGAGAAG
             228        0      +        0.18      CAGAAACGAG^AAGAACCTGA              228        0      +        0.18      CAGAAACGAG^AAGAACCTGA
             231        0      +        0.17      AAACGAGAAG^AACCTGACTG              231        0      +        0.17      AAACGAGAAG^AACCTGACTG
             371        2      +        0.25      TACCCCTCAG^CATGCTGATC              371        2      +        0.25      TACCCCTCAG^CATGCTGATC

Acceptor splice sites, complement strand                             Acceptor splice sites, complement strand
----------------------------------------                             ----------------------------------------
  pos 3'->5'  pos 5'->3'  phase strand  confidence 5'    intron exon   3'   pos 3'->5'  pos 5'->3'  phase strand  confidence 5'    intron exon   3'
    1391        170         2      -        0.25      TTTTAAACAG^TGGCATGAGG    1391        170         2      -        0.24      TTTTAAACAG^TGGCATGAGG
    1215        346         1      -        0.33      TTCTTTGCAG^TTAAAGGTAT    1215        346         1      -        0.33      TTCTTTGCAG^TTAAAGGTAT
     493       1068         0      -        0.07      CAGCCACCAG^GTGGCAACCT     493       1068         0      -        0.07      CAGCCACCAG^GTGGCAACCT
     469       1092         0      -        0.18      TATTGACCAG^ACACTTGGCC     469       1092         0      -        0.18      TATTGACCAG^ACACTTGGCC
     457       1104         2      -        0.18      ACTTGGCCAG^TTTGATGTTG     457       1104         2      -        0.18      ACTTGGCCAG^TTTGATGTTG
     438       1123         0      -        0.07      GGGCACCGAG^GACCTGAAGG     438       1123         0      -        0.07      GGGCACCGAG^GACCTGAAGG
     253       1308         2      -        0.27      TCACCCACAG^GCCTGTGTAA     253       1308         2      -        0.27      TCACCCACAG^GCCTGTGTAA
```

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.0.139.572497.0 :

Start	End	Score	Exon	Intron
266	280	0.78	gtgcccg**gt**atgacg	
517	531	0.83	ctggcag**gt**actgtg	
1103	1117	0.88	aattgag**gt**aatgta	
1134	1148	0.71	aatgatg**gt**tagcag	
1251	1265	1.00	atttcag**gt**aagggt	
1529	1543	0.98	atgacag**gt**tagtgt	

Donor site predictions for 10.42.1.119.572509.0 :

Start	End	Score	Exon	Intron
266	280	0.78	gtgcccg**gt**atgacg	
517	531	0.83	ctggcag**gt**actgtg	
1103	1117	0.88	aattgag**gt**aatgta	
1134	1148	0.71	aatgatg**gt**tagcag	
1251	1265	1.00	atttcag**gt**aagggt	
1529	1543	0.98	atgacag**gt**tagtgt	

Acceptor site predictions for 10.42.0.139.572497.0 :

Start	End	Score	Intron	Exon
40	80	0.41	ctcttgtgtgtcaccccct**ag**tctgactgacagtactccac	
152	192	0.78	cagtagcaggcctctttgc**ag**ggactctgcttcccaactgg	
230	270	0.49	agaacctgactgtttacac**ag**gcctgtgggtgaaatgtgcc	
334	374	0.75	ctggacctgcgtgtcctcc**ag**tttgccctacccctcagcat	
351	391	0.64	ccagtttgccctacccctc**ag**catgctgatcgccatgggtg	
414	454	0.83	aatgtgcaacactgccttc**ag**gtcctcggtgcccaacatca	
503	543	0.98	ggctgctattttcctggc**ag**gtactgtgagcctctcccca	
1515	1555	0.73	ttagaatgtgctttatgac**ag**gttagtgtttcctctgaggc	

Acceptor site predictions for 10.42.1.119.572509.0 :

Start	End	Score	Intron	Exon
40	80	0.41	ctcttgtgtgtcaccccct**ag**tctgactgacagtactccac	
152	192	0.78	cagtagcaggcctctttgc**ag**ggactctgcttcccaactgg	
230	270	0.49	agaacctgactgtttacac**ag**gcctgtgggtgaaatgtgcc	
334	374	0.75	ctggacctgcgtgtcctcc**ag**tttgccctacccctcagcat	
351	391	0.64	ccagtttgccctacccctc**ag**catgctgatcgccatgggtg	
414	454	0.83	aatgtgcaacactgccttc**ag**gtcctcggtgcccaacatca	
503	543	0.98	ggctgctattttcctggc**ag**gtactgtgagcctctcccca	
1515	1555	0.73	ttagaatgtgctttatgac**ag**gttagtgtttcctctgaggc	

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
aaacg(t/c)atttt	tatttt	catttt	31133	81%

CRYP-SKIP

Human Splicing Finder

Este tipo de nomenclatura no la permite.

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr	
wt	52	133	tgttgaaga	-1.00956740914	0.5625	4	14	30	0.2144376	
wt	52	127	agattactt	-1.88101073987	0.573770491803	1	11	26	0.029497937	
wt	52	114	ttgtgactt	1.10624650428	0.550458715596	1	16	32	1.2475164	
wt	52	99	acctcatgc	1.00201220116	0.5	3	9	14	0.89612524	
wt	52	87	tgtttaaaa	-3.9786620974	0.475609756098	12	6	13	-1.6409222	
wt	52	86	gtttaaaag	-1.44518207278	0.481481481481	11	6	13	-0.58375036	
wt	52	80	aagtaaaac	-0.527492388703	0.506666666667	5	6	13	0.16349253	
wt	52	69	attttaacg	-3.64777439711	0.5	18	10	16	-1.8553292	
wt	52	68	ttttaacga	0.714929736122	0.507936507937	17	10	16	-0.081262664	
wt	52	60	atgttagaa	-4.72411342716	0.509090909091	9	10	16	-1.7041458	
wt	52	54	gaataagac	-0.842110751682	0.551020408163	3	10	16	0.2091737	
wt	52	40	ttctaaata	0.246333338382	0.542857142857	4	12	17	0.57873205	
wt	52	34	atatcacct	-0.938217719104	0.551724137931	1	9	14	0.27973553	
wt	52	26	tacttatga	-1.01486748302	0.47619047619	21	0	0	-1.1710599	
wt	52	23	ttatgaata	-2.01321219191	0.5	18	0	0	-1.3643743	
wt	52	19	gaataacat	0.970303282875	0.571428571429	14	0	0	0.080075666	
mut	52	133	tgttgaaga	-1.00956740914	0.5625	4	14	30	0.2144376	
mut	52	127	agattactt	-1.88101073987	0.573770491803	1	11	26	0.029497937	
mut	52	114	ttgtgactt	1.10624650428	0.550458715596	1	16	32	1.2475164	
mut	52	99	acctcatgc	1.00201220116	0.5	3	9	14	0.89612524	
mut	52	87	tgtttaaaa	-3.9786620974	0.475609756098	36	10	16	-3.1321344	
mut	52	86	gtttaaaag	-1.44518207278	0.481481481481	35	10	16	-2.0749625	
mut	52	80	aagtaaaac	-0.527492388703	0.506666666667	29	10	16	-1.3277197	
mut	52	69	attttaacg	-3.64777439711	0.5	18	10	16	-1.8553292	
mut	52	68	ttttaacga	0.714929736122	0.507936507937	17	10	16	-0.081262664	
mut	52	60	atgttagaa	-4.72411342716	0.509090909091	9	10	16	-1.7041458	
mut	52	54	gaataagac	-0.842110751682	0.551020408163	3	10	16	0.2091737	
mut	52	40	ttctaaata	0.246333338382	0.542857142857	4	12	17	0.57873205	
mut	52	34	atatcacct	-0.938217719104	0.551724137931	1	9	14	0.27973553	
mut	52	26	tacttatga	-1.01486748302	0.47619047619	21	0	0	-1.1710599	
mut	52	23	ttatgaata	-2.01321219191	0.5	18	0	0	-1.3643743	
mut	52	19	gaataacat	0.970303282875	0.571428571429	14	0	0	0.080075666	

Variant Effect Predictor tool

Variant	Location	Allele	Consequence	Symbol	Gene	Feature type	Feature	Biotype	EXON	cDNA position						
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	3_prime_UTR_variant	CLDN12	ENSG00000157224	Transcript	ENST00000287916.8	protein_coding	3/3	1603	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	downstream_gene_variant	CLDN12	ENSG00000157224	Transcript	ENST00000394604.5	protein_coding	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	3_prime_UTR_variant	CLDN12	ENSG00000157224	Transcript	ENST00000394605.2	protein_coding	3/3	1526	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	downstream_gene_variant	CLDN12	ENSG00000157224	Transcript	ENST00000416322.5	protein_coding	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	downstream_gene_variant	CLDN12	ENSG00000157224	Transcript	ENST00000427904.1	protein_coding	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	intron_variant, non_coding_transcript_variant	CLDN12	ENSG00000157224	Transcript	ENST00000451941.6	processed_transcript	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	downstream_gene_variant	CLDN12	ENSG00000157224	Transcript	ENST00000462636.5	retained_intron	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	intron_variant, non_coding_transcript_variant	AC006153.1	ENSG00000273299	Transcript	ENST00000480135.1	lncRNA	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	intron_variant, non_coding_transcript_variant	CLDN12	ENSG00000157224	Transcript	ENST00000483862.5	processed_transcript	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	3_prime_UTR_variant	CLDN12	ENSG00000157224	Transcript	ENST00000496677.6	protein_coding	4/4	1571	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	intron_variant, non_coding_transcript_variant	CLDN12	ENSG00000157224	Transcript	ENST00000498033.5	processed_transcript	-	-	-	-	-	-	-	-
ENST00000394605.2:c.*581T>C	7:90413992-90413992	C	downstream_gene_variant	CLDN12	ENSG00000157224	Transcript	ENST00000498326.5	processed_transcript	-	-	-	-	-	-	-	-

ESEfinder

Se tienen dos resultados con sus acciones positivas sobre las materias 5':

| 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGTATTTTA | 3.40630 | | 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGTATTTTA | -11.32190 | | 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGTATTTTA | 2.79880 | | 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGTATTTTA | -14.57930 |
| 1406 (-155) | ACGTATTTTAACGATGTTAGAATAAGACTA | 0.97340 | | 1406 (-155) | ACGTATTTTAACGATGTTAGAATAAGACTA | -16.22440 | | 1406 (-155) | ACGTATTTTAACGATGTTAGAATAAGACTA | 0.57910 | | 1406 (-155) | ACGTATTTTAACGATGTTAGAATAAGACTA | -17.94610 |

Si se comparan con los resultados para la secuencia mutante se observa que las puntuaciones han aumentado ligeramente:

| 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGCATTTTA | 3.73950 | | 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGCATTTTA | -10.85380 | | 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGCATTTTA | 3.28730 | | 1386 (-175) | TGCCACTGTTTAAAAGTAAAACGCATTTTA | -14.15910 | | 1386 (-175) | TGCCACT | -4.00960 |
| 1406 (-155) | ACGCATTTTAACGATGTTAGAATAAGACTA | 1.74920 | | 1406 (-155) | ACGCATTTTAACGATGTTAGAATAAGACTA | -16.62950 | | 1406 (-155) | ACGCATTTTAACGATGTTAGAATAAGACTA | 1.32640 | | 1406 (-155) | ACGCATTTTAACGATGTTAGAATAAGACTA | -18.33060 | | 1406 (-155) | ACGCATT | -6.01670 |

Por lo tanto podría estar activándose o haciéndose más fuerte un sitio *donor*.

En cuento a los sitios ESE, parece que algunos de lo sitios se están modificando, por lo que podrían estar afectando al *splicing*:

Pos	Seq	Value
1403 (-158)	AAAACGT	1.42065
1404 (-157)	AAACGTA	-1.94380
1405 (-156)	AACGTAT	-4.27938
1406 (-155)	ACGTATT	-5.33565
1407 (-154)	CGTATTT	-1.16515
1408 (-153)	GTATTTT	-5.86862
1409 (-152)	TATTTTA	-5.76604

Pos	Seq	Value
1403 (-158)	AAAACGT	0.27244
1404 (-157)	AAACGTA	-1.65690
1405 (-156)	AACGTAT	-3.27373
1406 (-155)	ACGTATT	-4.14202
1407 (-154)	CGTATTT	0.04748
1408 (-153)	GTATTTT	-4.93691
1409 (-152)	TATTTTA	-4.78866

Pos	Seq	Value
1403 (-158)	AAAACGTA	-0.73529
1404 (-157)	AAACGTAT	-2.58192
1405 (-156)	AACGTATT	-2.65561
1406 (-155)	ACGTATTT	-4.33651
1407 (-154)	CGTATTTT	-3.59235
1408 (-153)	GTATTTTA	1.62430
1409 (-152)	TATTTTAA	-0.71987

Pos	Seq	Value
1403 (-158)	AAAACGT	-2.67967
1404 (-157)	AAACGTA	-3.95077
1405 (-156)	AACGTAT	-4.30710
1406 (-155)	ACGTATT	-4.12293
1407 (-154)	CGTATTT	-2.12015
1408 (-153)	GTATTTT	-3.52663
1409 (-152)	TATTTTA	-4.02423

Pos	Seq	Value
1403 (-158)	AAAACGC	-0.43275
1404 (-157)	AAACGCA	-2.40151
1405 (-156)	AACGCAT	-1.75565
1406 (-155)	ACGCATT	-3.87833
1407 (-154)	CGCATTT	1.14487
1408 (-153)	GCATTTT	-6.47155
1409 (-152)	CATTTTA	-2.81555

Pos	Seq	Value
1403 (-158)	AAAACGC	-0.92425
1404 (-157)	AAACGCA	-1.76090
1405 (-156)	AACGCAT	-1.46194
1406 (-155)	ACGCATT	-2.88726
1407 (-154)	CGCATTT	2.11634
1408 (-153)	GCATTTT	-5.18569
1409 (-152)	CATTTTA	-1.68361

Pos	Seq	Value
1403 (-158)	AAAACGCA	-1.37973
1404 (-157)	AAACGCAT	-0.85291
1405 (-156)	AACGCATT	-1.89637
1406 (-155)	ACGCATTT	-4.10567
1407 (-154)	CGCATTTT	-3.02342
1408 (-153)	GCATTTTA	0.23590
1409 (-152)	CATTTTAA	-0.69874

Pos	Seq	Value
1403 (-158)	AAAACGC	-1.14057
1404 (-157)	AAACGCA	-2.23385
1405 (-156)	AACGCAT	-5.79835
1406 (-155)	ACGCATT	-1.73830
1407 (-154)	CGCATTT	-3.46842
1408 (-153)	GCATTTT	-3.21458
1409 (-152)	CATTTTA	-4.38092

EX-SKIP

No se puede analizar porque las secuencias son demasiado largas.

HOT-SKIP

ttcctgctctgttctgctattgtcccctcatgatttgtcctcttgtgtgtcaccccctagTCTGACTGACAGTACTCCAC
AAGCTTGCCTGCCATGGGCTGTCGGGATGTCCACGCAGCCACAGTCCTTTCCTTCCTGTGTGGAATCGCCTCAGTAGCAG
GCCTCTTTGCAGGGACTCTGCTTCCCAACTGGAGAAAATTACGATTGATCACATTCAACAGAAACGAGAAGAACCTGACT
GTTTACACAGGCCTGTGGGTGAAATGTGCCCGGTATGACGGGAGCAGTGACTGCCTGATGTACGACACTACTTGGTACTC
ATCAGTTGACCAGCTGGACCTGCGTGTCCTCCAGTTTGCCCTACCCCTCAGCATGCTGATCGCCATGGGTGCCCTGCTGC
TCTGCCTGATTGGAATGTGCAACACTGCCTTCAGGTCCTCGGTGCCCAACATCAAACTGGCCAAGTGTCTGGTCAATAGT
GCAGGTTGCCACCTGGTGGCTGGGCTGCTATTTTTCCTGGCAGGTACTGTGAGCCTCTCCCCATCTATCTGGGTCATCTT
TTATAACATCCATCTGAACAAGAAGTTTGAGCCAGTCTTTTCATTTGACTATGCAGTGTATGTCACTATTGCTAGTGCTG
GGGGCCTGTTTATGACTTCCCTTATACTATTTATTTGGTATTGTACATGCAAATCTTTGCCTTCTCCTTTCTGGCAACCA
TTGTACTCCCATCCACCCAGTATGCATACTTACTCACAGCCCTATTCAGCACGCTCTCGCCTCTCTGCCATTGAAATTGA
CATTCCAGTAGTTTCACACACCACTTAATGGGGAAATAGTTAATTGTTAAAGAAAACTTCTTGTAGCCTCACATTCCCCT
TGTGCAAAGAGCTCTTTTGGACCTACATACATTTTCCTTTGTTTTTGACCAATCAATGAAGCCAAATTTATATGTCCTAG
TAGAATGAAGTGCTGCTAGTTTTTATGAGAAGTATATTATATTAAATGTGAATTTTTTAAATTTTGCTTCTTATACTGGA
AGGAATTTTAGCCTTCATATTGATATCTAATTAATTATTTAAGTGGAAGAGGCCTGCATCACAATTGAGGTAATGTAGAG
CAACATGTTAAAGAATGATGGTTAGCAGAAGCTGTTGTATACAATCTTCATGAAAATTTCAGTGTGTATTTTTCTTTTTC
TATAATACCTTTAACTGCAAAGAAAAGGCAGTTTCAAATATAAGAAATTTATTTCAGGTAAGGGTAATATTTTAATAGTA
GTCAATAATCTAGCTTAAGGCTGTAACTCTTCTATCGGGGCTAATTGTATGAATAGGTGTCAGTATGTTGAAGATTACTT
TCTTTTGTGACTTTCTTCTACCTCATGCCACTGTTTAAAAGTAAAACGTATTTTAACGATGTTAGAATAAGACTACCATT
CTAAATATCACCTACTTATGAATAACATGTAATAATTTTTAACATTAATGATTCCATAAATTGTATTATTGGGATTAGAA
TGTGCTTTATGACAGGTTAGTGTTTCCTCTGAGGCAGAAAACTCTTTTTTGGAGATATCTTCCATCAAGCAGTACTCGTG
CCCATATACAATCTCTTAGTGGCTAGGAGAAATAAATAAAAGGGCCATAATGGTTTGTTCTCTTTCAGACATAATTTAGT
AGGGGACAAGAAGTCTGTTCTTCAGTGAGTACACTAGAGATTTACTCTGGTGACTGCCTTTTGAGTTATGGGTGAAGTAA
GGTATGGCTTTACCATAACCTTGATTCATTCACCCTTGATTCATTTCTCGCCCCCGTCACTGATTATTTCCTTGAGCATA
TATCTCTGCCTAACACTTTAGTAGGTGCTATAGAGGATACATGAAAAGTATGAGATCTGGTTCCATCCAGTAAGACATTT
TAATAGAGAAGATCAAAATGTTACCTGGCAGTTGGGGAATAATCTGACTTCGTTGGCAGTTGGCCTTAACTTCTTAATCA
TTGATCCAGGAATATTTCAACCAGAGACACAACTTTCTGGCAGACAGACAAATTGTACAACACCAACAATATCCTGGACC
TTGAAATTCTGTTTACTTCAGTCCATTGTATCCTTTAAGGCACCTGTGCTAGCCTAGATTTTGTAATAACACTGATTTAT
GAGAATGGACAAAAGTGGTAGGGAAATTGTTCCCTCTCCACTTCTGAAAGTATGATGATGTATTAAGGATGGAGGAGTTA
TTAAAAATGTCTCTTCTGATGAGGTAACAATTAGATGAAACCATGTTAAAGCTGAGATGAACACTTAGAAATTCAGGGAT
ATTGGGTCTTTAGCCTTATGAATTTGAGCTGCTTATTTAATTGGTGTAATTTACTACATATTAGTACTATATTCGTAAGG
ATTTTTTATTAACCATTACAGATTTTACAAACAGCTAGTTATATGGTAAACAGATTATTATGCCTTTTTGCAATTCTGAA
TATGATTCTAGTATTTGTGTAGATGTATTTGGTACTTTTTCCCCTAATTCCAACACTAGTTTATATATATAGCGAATAAA
TCTAGTTGTATAAATTTTTAAATGCCGTCAGTAGAAAGCACACAAGGTTATGATTTTTTTAATTACTGGCTTCTGATTTC
TTTCACTTCTGATCCTTTTCCTTTTTCTCAGATGTAGCTGAGTCTTGATCATTTTAAGACAACGATGGGTAGAATTTTGA
GATTAATGTTAATTTTCCCTTTTTGTTAATTTCAGTCCCCTCTCACTATGCTTTTGTCCAGAAGGATCAAGAATTCTACC
ATCCCTTGGGTCTTTGTGTATAAACAATGTTAAATAAAGGTAGACTCAGTCTTTAAGATATTAGACAGTTTTTTTAGTCC
ATGGGATTGTAAATATAAACATTAACTTTCCTATAAGAATATTTTGGCTTTGTAATCTATAGCCTCAAATTGGTATTTAT
TATGGATTCACTAGACAAACAGCTGTTTCCTTATTGTCTTTTTTCTTTAGTGTTTCTGATTTGCTATCAGTAGCTGTTTT
TAAAGCCGTCCAAGGAAAATAATTATTTTACAGTTTTTGAAGTCACTTTTGAGCCCTCATCAAGCTCTCATTGTGATGGGA
GGGATACCTTTTTTGTTGTTAAAAGCCTATTATTGTTAAAGGCCTTTTATGGAAACCAACTTGGAAAACAACCTTAAATGT
GGATGTATCAGATTTGGTTTATCCAGCCATGGGAGAGAAAACAAACCTAAGTTTACTTTACTTGTACATATACACTACAA
TGGATAGTATATTTGCTGTAAACTACAATGTAAAACCTCAATAAAAGTGCGCTGTACTTCTTAATGTTTATTAAAAGATG
TATTTTTACAAgtttctgcttctgtttttttcctcagtgtttttaaggtgaaatattttttgagaaatgtttct