Ejemplo comparación de resultados predictores in sillico
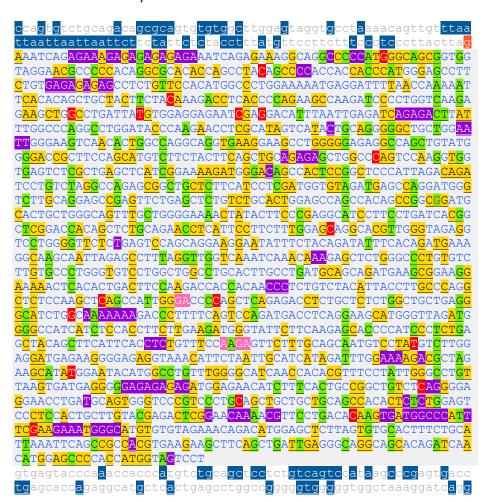
Cambio de estudio NLRP1 c.1308G>A (chr17:5559388 G/A, rs200867395 o NM_033004.4: c.1308G>A)

Exón 4 e intrones adyacentes:

ccagtgtctgcagacagcgcagtgtgtggcttggagtaggtgcctaaaacagttgtttaa
ttaattaattaattcttctattctctacctttatgttccttctttccctccccttacttag
AAATCAGAGAAAGAGAGAGAGAGAAATCAGAGAAAGGCAGGCCCCCATGGGCAGCGGTGG
TAGGAACGCCCCCACAGGCGCACACCAGCCTACAGCCCCACCACCACCCATGGGAGCCTT
CTGTGAGAGAGAGCCTCTGTTCCACATGGCCCTGGAAAAATGAGGATTTTAACCAAAAAT
TCACACAGCTGCTACTTCTACAAAGACCTCACCCCAGAAGCCAAGATCCCCTGGTCAAGA
GAAGCTGGCCTGATTATGTGGAGGAGAATCGAGGACATTTAATTGAGATCAGAGACTTAT
TTGGCCCAGGCCTGGATACCCAAGAACCTCGCATAGTCATACTGCAGGGGGCTGCTGGAA
TTGGGAAGTCAACACTGGCCAGGCAGGTGAAGGAAGCCTGGGGGGAGAGGCCAGCTGTATG
GGGACCGCTTCCAGCATGTCTTCTACTTCAGCTGCAGAGAGCTGGCCCAGTCCAAGGTGG
TGAGTCTCGCTGAGCTCATCGGAAAAGATGGGACAGCCACTCCGGCTCCCATTAGACAGA
TCCTGTCTAGGCCAGAGCGGCTGCTCTTCATCCTCGATGGTGTAGATGAGCCAGGATGGG
TCTTGCAGGAGCCGAGTTCTGAGCTCTGTCTGCACTGGAGCCAGCCACAGCCGGCGGATG
CACTGCTGGGCAGTTTGCTGGGGAAAACTATACTTCCCGAGGCATCCTTCCTGATCACGG
CTCGGACCACAGCTCTGCAGAACCTCATTCCTTCTTTGGAGCAGGCACGTTGGGTAGAGG
TCCTGGGGTTCTCTGAGTCCAGCAGGAAGGAATATTTCTACAGATATTTCACAGATGAAA
GGCAAGCAATTAGAGCCTTTAGGTTGGTCAAATCAAACAAAGAGCTCTGGGCCCTGTGTC
TTGTGCCCTGGGTGTCCTGGCTGGCCTGCACTTGCCTGATGCAGCAGATGAAGCGGAAGG
AAAAACTCACACTGACTTCCAAGACCACCACAACCCTCTGTCTACATTACCTTGCCCAGG
CTCTCCAAGCTCAGCCATTGGGACCCCAGCTCAGAGACCTCTGCTCTCTGGCTGCTGAGG
GCATCTGCAAAAAAAGACCCTTTTCAGTCCAGATGACCTCAGGAAGCATGGGTTAGATG
GGGCCATCATCTCCACCTTCTTGAAGATGGGTATTCTTCAAGAGCACCCCATCCCTCTGA
GCTACAGCTTCATTCACCTCTGTTTCCAAGAGTTCTTTGCAGCAATGTCCTATGTCTTGG
AGGATGAGAAGGGGGAGAGGTAAACATTCTAATTGCATCATAGATTTGGAAAAGACGCTAG
AAGCATATGGAATACATGGCCTGTTTGGGGCATCAACCACACGTTTCCTATTGGGCCTGT
TAAGTGATGAGGGGGAGAGAGAGATGGAGAACATCTTTCACTGCCGGCTGTCTCAGGGGA
GGAACCTGATGCAGTGGGTCCCGTCCCTGCAGCTGCTGCTGCAGCCACACTCTCTGGAGT
CCCTCCACTGCTTGTACGAGACTCGGAACCAAAACGTTCCTGACACAAGTGATGGCCCATT
TCGAAGAAATGGGCATGTGTGTAGAAACAGACATGGAGCTCTTAGTGTGCACTTTCTGCA
TTAAATTCAGCCGCCACGTGAAGAAGCTTCAGCTGATTGAGGGCAGGCAGCACAGATCAA
CATGGAGCCCACCATGGTAGTCCT

gtgagtacccaaaccacccatgtctgcagctcctctgtcagtccataagcccgagtgacc
tgagcaccagaggcatgctcactgagcctggccggggggtggggtggctaaaggatcagg

El cambio se encuentra en línea 11 del exón 4 (la **g** en color rojo subrayado en verde).

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

NetGene2

Acceptor splice sites, direct strand

```
     pos 5'->3'  phase strand  confidence  5'     intron exon      3'
        120        1      +       0.77       CCTTACTTAG^AAATCAGAGA
        325        1      +       0.14       TTCTACAAAG^ACCTCACCCC
        337        2      +       0.14       CTCACCCCAG^AAGCCAAGAT
        444        1      +       0.07       GATACCCAAG^AACCTCGCAT
        533        0      +       0.19       GAGAGGCCAG^CTGTATGGGG
        554        0      +       0.18       CCGCTTCCAG^CATGTCTTCT
        571        2      +       0.67       TCTACTTCAG^CTGCAGAGAG
        577        2      +       0.25       TCAGCTGCAG^AGAGCTGGCC
        659        0      +       0.14       CATTAGACAG^ATCCTGTCTA
        760        0      +       0.07       TGCACTGGAG^CCAGCCACAG
        764        1      +       0.17       CTGGAGCCAG^CCACAGCCGG
        770        1      +       0.17       CCAGCCACAG^CCGGCGGATG
       1064        0      +       0.30       CCTGATGCAG^CAGATGAAGC
       1067        0      +       0.19       GATGCAGCAG^ATGAAGCGGA
       1199        0      +       0.14       GGCTGCTGAG^GGCATCTGGC
       1217        0      +       0.07       GCAAAAAAAG^ACCCTTTTCA
       1228        2      +       0.18       CCCTTTTCAG^TCCAGATGAC
       1233        1      +       0.47       TTCAGTCCAG^ATGACCTCAG
       1243        2      +       0.18       ATGACCTCAG^GAAGCATGGG
       1247        0      +       0.17       CCTCAGGAAG^CATGGGTTAG
       1257        1      +       0.07       CATGGGTTAG^ATGGGGCCAT
       1511        0      +       0.07       AAGTGATGAG^GGGGAGAGAG
       1556        0      +       0.42       GCTGTCTCAG^GGGAGGAACC
       1561        2      +       0.19       CTCAGGGGAG^GAACCTGATG
       1574        0      +       0.19       CCTGATGCAG^TGGGTCCCGT
       1592        0      +       0.18       GTCCCTGCAG^CTGCTGCTGC
```

Acceptor splice sites, direct strand

```
     pos 5'->3'  phase strand  confidence  5'     intron exon      3'
        120        1      +       0.77       CCTTACTTAG^AAATCAGAGA
        325        1      +       0.14       TTCTACAAAG^ACCTCACCCC
        337        2      +       0.14       CTCACCCCAG^AAGCCAAGAT
        444        1      +       0.07       GATACCCAAG^AACCTCGCAT
        533        0      +       0.19       GAGAGGCCAG^CTGTATGGGG
        554        0      +       0.18       CCGCTTCCAG^CATGTCTTCT
        571        2      +       0.67       TCTACTTCAG^CTGCAGAGAG
        577        2      +       0.25       TCAGCTGCAG^AGAGCTGGCC
        659        0      +       0.14       CATTAGACAG^ATCCTGTCTA
        764        1      +       0.07       CTGGAGCCAG^CCACAGCCGG
        770        1      +       0.07       CCAGCCACAG^CCGGCAGATG
        777        2      +       0.17       CAGCCGGCAG^ATGCACTGCT
        793        2      +       0.07       TGCTGGGCAG^TTTGCTGGGG
       1064        0      +       0.30       CCTGATGCAG^CAGATGAAGC
       1067        0      +       0.19       GATGCAGCAG^ATGAAGCGGA
       1199        0      +       0.14       GGCTGCTGAG^GGCATCTGGC
       1217        0      +       0.07       GCAAAAAAAG^ACCCTTTTCA
       1228        2      +       0.18       CCCTTTTCAG^TCCAGATGAC
       1233        1      +       0.47       TTCAGTCCAG^ATGACCTCAG
       1243        2      +       0.18       ATGACCTCAG^GAAGCATGGG
       1247        0      +       0.17       CCTCAGGAAG^CATGGGTTAG
       1257        1      +       0.07       CATGGGTTAG^ATGGGGCCAT
       1511        0      +       0.07       AAGTGATGAG^GGGGAGAGAG
       1556        0      +       0.42       GCTGTCTCAG^GGGAGGAACC
       1561        2      +       0.19       CTCAGGGGAG^GAACCTGATG
       1574        0      +       0.19       CCTGATGCAG^TGGGTCCCGT
       1592        0      +       0.18       GTCCCTGCAG^CTGCTGCTGC
```

Los cambios entre ambas se encuentran en los sitios *acceptor*. Desaparece uno de los sitios de la secuencia WT a la mutante (en rojo), uno se ve alterado por la presencia de la mutación (en verde) y aparecen dos sitios nuevos en la secuencia mutante (en azul). Si alguno de estos fuera detectado por el *spliceosome* en vez del *acceptor* normal del exón, se perderían los primeros 655 o 673 nucleótidos del exón.

Splice Site Prediction by Neural Network (NNSplice)

Donor site predictions for 10.42.1.119.573119.0 :

Start	End	Score	Exon	Intron
593	607	0.99	caaggtg**gt**gagtct	
976	990	0.53	cctttag**gt**tggtca	
1392	1406	0.76	gggagag**gt**aaacat	
1819	1833	0.95	tagtcct**gt**gagtac	

Donor site predictions for 10.42.2.148.573131.0 :

Start	End	Score	Exon	Intron
593	607	0.99	caaggtg**gt**gagtct	
976	990	0.53	cctttag**gt**tggtca	
1392	1406	0.76	gggagag**gt**aaacat	
1819	1833	0.95	tagtcct**gt**gagtac	

Acceptor site predictions for 10.42.1.119.573119.0 :

Start	End	Score	Intron	Exon
100	140	0.82	ttctttccctcccttactt**ag**aaatcagagaaagagagaga	
551	591	0.80	ccagcatgtcttctacttc**ag**ctgcagagagctggcccagt	
639	679	0.42	actccggctcccattagac**ag**atcctgtctaggccagagcg	
864	904	0.72	ctcattccttctttggagc**ag**gcacgttgggtagaggtcct	
923	963	0.50	caggaaggaatatttctac**ag**atatttcacagatgaaaggc	
1119	1159	0.65	tgtctacattaccttgccc**ag**gctctccaagctcagccatt	
1266	1306	0.53	atcatctccaccttcttga**ag**atgggtattcttcaagagca	
1402	1442	0.53	aacattctaattgcatcat**ag**atttggaaaagacgctagaa	
1846	1886	0.43	tgtctgcagctcctctgtc**ag**tccataagcccgagtgacct	

Acceptor site predictions for 10.42.2.148.573131.0 :

Start	End	Score	Intron	Exon
100	140	0.82	ttctttccctcccttactt**ag**aaatcagagaaagagagaga	
551	591	0.80	ccagcatgtcttctacttc**ag**ctgcagagagctggcccagt	
639	679	0.42	actccggctcccattagac**ag**atcctgtctaggccagagcg	
864	904	0.72	ctcattccttctttggagc**ag**gcacgttgggtagaggtcct	
923	963	0.50	caggaaggaatatttctac**ag**atatttcacagatgaaaggc	
1119	1159	0.65	tgtctacattaccttgccc**ag**gctctccaagctcagccatt	
1266	1306	0.53	atcatctccaccttcttga**ag**atgggtattcttcaagagca	
1402	1442	0.53	aacattctaattgcatcat**ag**atttggaaaagacgctagaa	
1846	1886	0.43	tgtctgcagctcctctgtc**ag**tccataagcccgagtgacct	

Spliceman

Point mutation	Wildtype (wt)	Mutation (mt)	L1 distance	Ranking (L1)
ccggc(g/a)gatgc	cggatg	cagatg	29871	73%

CRYP-SKIP

Human Splicing Finder

New Acceptor splice site	Activation of a cryptic Acceptor site. Potential alteration of splicing (cryptic exon activation)

Algorithm/Matix	position	sequences	variation
HSF Acceptor site (matrix AG)	chr17:5559398	- REF : CACAGCCGGCGGAT - ALT : CACAGCCGGCAGAT	46.21 > 74.08 => 60.31%

Se perderían los primeros 657 nt del exón.

SVM-BPfinder

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
wt	28	108	ttctgagct		1.27304608698	0.52427184466	2	10	16	1.0920169
wt	28	16	tcctgatca		1.38985199703	0.545454545455	11	0	0	0.42585494
mut	28	108	ttctgagct		1.27304608698	0.52427184466	2	10	16	1.0920169
mut	28	16	tcctgatca		1.38985199703	0.545454545455	11	0	0	0.42585494

Variant Effect Predictor tool

ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000262467.10	protein_coding	4/16	1882	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000269280.8	protein_coding	5/17	1466	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000345221.7	protein_coding	5/17	1674	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000354411.7	protein_coding	4/16	1308	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant, NMD_transcript_variant	NLRP1	ENSG00000091592	Transcript	ENST00000544378.6	nonsense_mediated_decay	4/17	1860	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	non_coding_transcript_exon_variant	NLRP1	ENSG00000091592	Transcript	ENST00000571307.1	processed_transcript	5/14	1674	-	-	-	-	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000571451.6	protein_coding	4/16	1830	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000572272.6	protein_coding	4/17	1860	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000577119.5	protein_coding	4/15	1308	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000613500.4	protein_coding	5/17	1674	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000617618.4	protein_coding	5/18	1674	1308	436	A	GCG/GCA	rs200867395, COSV52574053
ENST00000354411.7:c.1308G>A	17:5559388-5559388	T	synonymous_variant	NLRP1	ENSG00000091592	Transcript	ENST00000619223.4	protein_coding	5/17	1674	1308	436	A	GCG/GCA	rs200867395, COSV52574053

ESEfinder

Las predicciones que cambian de positivo a negativo (o viceversa) entre secuencias son 750, 756, 759 y 763:

750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCGGAT	-10.57380	750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCGGAT	1.57560	750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCGGAT	-10.87090	750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCGGAT	1.18890
756 (-1190)	TGGAGCCAGCCACAGCCGGCGGATGCACTG	-10.85340	756 (-1190)	TGGAGCCAGCCACAGCCGGCGGATGCACTG	2.04830	756 (-1190)	TGGAGCCAGCCACAGCCGGCGGATGCACTG	-11.19400	756 (-1190)	TGGAGCCAGCCACAGCCGGCGGATGCACTG	1.90650
759 (-1187)	AGCCAGCCACAGCCGGCGGATGCACTGCTG	0.04810	759 (-1187)	AGCCAGCCACAGCCGGCGGATGCACTGCTG	-14.23440	759 (-1187)	AGCCAGCCACAGCCGGCGGATGCACTGCTG	-0.48790	759 (-1187)	AGCCAGCCACAGCCGGCGGATGCACTGCTG	-12.41270
763 (-1183)	AGCCACAGCCGGCGGATGCACTGCTGGGCA	-6.82740	763 (-1183)	AGCCACAGCCGGCGGATGCACTGCTGGGCA	-15.33490	763 (-1183)	AGCCACAGCCGGCGGATGCACTGCTGGGCA	-7.21720	763 (-1183)	AGCCACAGCCGGCGGATGCACTGCTGGGCA	-13.57530

750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCAGAT	-10.98490
756 (-1190)	TGGAGCCAGCCACAGCCGGCAGATGCACTG	-11.16740
759 (-1187)	AGCCAGCCACAGCCGGCAGATGCACTGCTG	0.73660
763 (-1183)	AGCCACAGCCGGCAGATGCACTGCTGGGCA	-4.52190

750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCAGAT	1.24240
756 (-1190)	TGGAGCCAGCCACAGCCGGCAGATGCACTG	1.70610
759 (-1187)	AGCCAGCCACAGCCGGCAGATGCACTGCTG	-14.42760
763 (-1183)	AGCCACAGCCGGCAGATGCACTGCTGGGCA	1.24440

750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCAGAT	-11.40570
756 (-1190)	TGGAGCCAGCCACAGCCGGCAGATGCACTG	-11.58870
759 (-1187)	AGCCAGCCACAGCCGGCAGATGCACTGCTG	0.18750
763 (-1183)	AGCCACAGCCGGCAGATGCACTGCTGGGCA	-4.85460

750 (-1196)	CTGCACTGGAGCCAGCCACAGCCGGCAGAT	0.85250
756 (-1190)	TGGAGCCAGCCACAGCCGGCAGATGCACTG	1.63460
759 (-1187)	AGCCAGCCACAGCCGGCAGATGCACTGCTG	-12.54050
763 (-1183)	AGCCACAGCCGGCAGATGCACTGCTGGGCA	0.96200

Todas aumentan sus puntuaciones, lo que nos indica que se está activando un sitio *acceptor*.

Existen algunos cambios destacados en las matrices de ESE:

770 (-1176)	GCCGGCG	-3.53089
771 (-1175)	CCGGCGG	0.98204
772 (-1174)	CGGCGGA	4.29797
773 (-1173)	GGCGGAT	-1.87911
774 (-1172)	GCGGATG	-5.23902
775 (-1171)	CGGATGC	1.16527
776 (-1170)	GGATGCA	-3.37818

770 (-1176)	GCCGGCG	-2.18768
771 (-1175)	CCGGCGG	2.18449
772 (-1174)	CGGCGGA	4.70413
773 (-1173)	GGCGGAT	-1.13409
774 (-1172)	GCGGATG	-4.17133
775 (-1171)	CGGATGC	2.08168
776 (-1170)	GGATGCA	-2.47988

770 (-1176)	GCCGGCGG	-1.50025
771 (-1175)	CCGGCGGA	-6.30386
772 (-1174)	CGGCGGAT	-2.98070
773 (-1173)	GGCGGATG	1.57904
774 (-1172)	GCGGATGC	-7.37174
775 (-1171)	CGGATGCA	-3.35146
776 (-1170)	GGATGCAC	1.02918

770 (-1176)	GCCGGCG	-3.15999
771 (-1175)	CCGGCGG	-0.10239
772 (-1174)	CGGCGGA	-2.39893
773 (-1173)	GGCGGAT	-6.62548
774 (-1172)	GCGGATG	-2.53560
775 (-1171)	CGGATGC	-0.12002
776 (-1170)	GGATGCA	-4.62803

770 (−1176)	GCCGGCA	−2.80410
771 (−1175)	CCGGCAG	−1.59569
772 (−1174)	CGGCAGA	2.37836
773 (−1173)	GGCAGAT	1.02947
774 (−1172)	GCAGATG	−7.30280
775 (−1171)	CAGATGC	1.61249
776 (−1170)	AGATGCA	−4.30606

770 (−1176)	GCCGGCA	−1.75231
771 (−1175)	CCGGCAG	0.40483
772 (−1174)	CGGCAGA	2.99774
773 (−1173)	GGCAGAT	0.58566
774 (−1172)	GCAGATG	−5.55509
775 (−1171)	CAGATGC	1.78346
776 (−1170)	AGATGCA	−2.71743

770 (−1176)	GCCGGCAG	−0.20335
771 (−1175)	CCGGCAGA	−6.66535
772 (−1174)	CGGCAGAT	−2.55543
773 (−1173)	GGCAGATG	1.57904
774 (−1172)	GCAGATGC	−6.07485
775 (−1171)	CAGATGCA	−3.71231
776 (−1170)	AGATGCAC	−0.72289

770 (−1176)	GCCGGCA	−5.54173
771 (−1175)	CCGGCAG	−0.67022
772 (−1174)	CGGCAGA	0.15563
773 (−1173)	GGCAGAT	−6.47018
774 (−1172)	GCAGATG	0.07319
775 (−1171)	CAGATGC	−1.56519
776 (−1170)	AGATGCA	−3.17331

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	16	79	52	343	4923.0101	170	-238.8054	114	174	595	7673.1704	655	879.0698	660	1538	0.43
mut	16	79	52	343	4923.0101	170	-238.8054	114	175	597	7692.7496	658	881.4372	660	1544	0.43

Both alleles have a comparable chance of exon skipping.

HOT-SKIP

ccagtgtctgcagacagcgcagtgtgtggcttggagtaggtgcctaaaacagttgtttaattaattaattaattcttcta
ttctctacctttatgttccttctttccctcccttacttagAAATCAGAGAAAGAGAGAGAGAGAAATCAGAGAAAGGCAG
GCCCCCATGGGCAGCGGTGGTAGGAACGCCCCCACAGGCGCACACCAGCCTACAGCCCCACCACCACCCATGGGAGCCTT
CTGTGAGAGAGAGCCTCTGTTCCACATGGCCCTGGAAAAATGAGGATTTTAACCAAAAATTCACACAGCTGCTACTTCTA
CAAAGACCTCACCCCAGAAGCCAAGATCCCCTGGTCAAGAGAAGCTGGCCTGATTATGTGGAGGAGAATCGAGGACATTT
AATTGAGATCAGAGACTTATTTGGCCCAGGCCTGGATACCCAAGAACCTCGCATAGTCATACTGCAGGGGGGCTGCTGGAA
TTGGGAAGTCAACACTGGCCAGGCAGGTGAAGGAAGCCTGGGGGAGAGGCCAGCTGTATGGGGACCGCCTTCCAGCATGTC
TTCTACTTCAGCTGCAGAGAGCTGGCCCAGTCCAAGGTGGTGAGTCTCGCTGAGCTCATCGGAAAAGATGGGACAGCCAC
TCCGGCTCCCATTAGACAGATCCTGTCTAGGCCAGAGCGGCTGCTCTTCATCCTCGATGGTGTAGATGAGCCAGGATGGG
TCTTGCAGGAGCCGAGTTCTGAGCTCTGTCTGCACTGGAGCCAGCCACAGCCGGCGGATGCACTGCTGGGCAGTTTGCTG
GGGAAAACTATACTTCCCGAGGCATCCTTCCTGATCACGGCTCGGACCACAGCTCTGCAGAACCTCATTCCTTCTTTGGA
GCAGGCACGTTGGGTAGAGGTCCTGGGGTTCTCTGAGTCCAGCAGGAAGGAATATTTCTACAGATATTTCACAGATGAAA
GGCAAGCAATTAGAGCCTTTAGGTTGGTCAAATCAAACAAAGAGCTCTGGGCCCTGTGTCTTGTGCCCTGGGTGTCCTGG
CTGGCCTGCACTTGCCTGATGCAGCAGATGAAGCGGAAGGAAAAACTCACACTGACTTCCAAGACCACCACAACCCTCTG
TCTACATTACCTTGCCCAGGCTCTCCAAGCTCAGCCATTGGGACCCCAGCTCAGAGACCTCTGCTCTCTGGCTGCTGAGG
GCATCGGCAAAAAAAGACCCTTTTCAGTCCAGATGACCTCAGGAAGCATGGGTTAGATGGGGCCATCATCTCCACCTTC
TTGAAGATGGGTATTCTTCAAGAGCACCCCATCCCTCTGAGCTACAGCTTCATTCACCTCTGTTTCCAAGAGTTCTTTGC
AGCAATGTCCTATGTCTTGGAGGATGAGAAGGGGAGAGGTAAACATTCTAATTGCATCATAGATTTGGAAAAGACGCTAG
AAGCATATGGAATACATGGCCTGTTTGGGGCATCAACCACACGTTTCCTATTGGGCCTGTTAAGTGATGAGGGGGAGAGA
GAGATGGGAACATCTTTCACTGCCGGCTGTCTCAGGGGAGGAACCTGATGCAGTGGGTCCCGTCCCTGCAGCTGCTGCT
GCAGCCACACTCTCTGGAGTCCCTCCACTGCTTGTACGAGACTCGGAACAAAACGTTCCTGACACAAGTGATGGCCCATT
TCGAAGAAATGGGCATGTGTGTAGAAACAGACATGGAGCTCTTAGTGTGCACTTTCTGCATTAAATTCAGCCGCCACGTG
AAGAAGCTTCAGCTGATTGAGGGCAGGCAGCACAGATCAACATGGAGCCCCACCATGGTAGTCCTgtgagtacccaaacc
acccatgtctgcagctcctctgtcagtccataagcccgagtgacctgagcaccagaggcatgctcactgagcctggccgg
gggtgggggtggctaaaggatcagg