

## SUPPLEMENTARY MATERIAL

### Asalign: multiple alignment tool for revealing alternative splicing patterns

Multiple alignment is an important way of revealing alternative splicing patterns of eukaryotic genes. In order to give a clear and precise alignment of a group of alternatively spliced mRNAs, we have tried some popular multiple alignment tools, including CLUSTALW (1) and DIALIGN (2). These tools, however, turned out to be unsuitable for our purpose because of being time-consuming and resulting in strong noise in some results. Thus a new tool to align mRNAs, DNAs and ESTs of alternatively spliced genes was developed. This tool should give a precise and clear result in a short time. The effort resulted in Asalign, the tool introduced here.

Asalign is a heuristic algorithm, which guides segment-to-segment alignment along a guide-tree. The idea of segment-to-segment alignment was introduced by Morgenstern for multiple alignment. We have modified the way to realize the idea. Figure S1 illustrates the main steps of the alignment.

Three mRNAs  $s_1$ ,  $s_2$ ,  $s_3$  are shown in Figure S1a. Assume that they are isoforms spliced in different ways from one pre-mRNA  $s_0$ , and each mRNA represents a node whose index is 1, 2 or 3. In order to align  $s_1$ ,  $s_2$ ,  $s_3$ , a guide-tree is created as shown in Figure S1b. Alignment is carried out between two sequences each time. At first,  $s_1$  and  $s_2$  will be aligned. The optimal group of their consistent segments will be picked out. Here, the selected segments are A and C. According to the consistent segments, a topological structure can be obtained, which records the relationships among all segments (refer to Fig. S1c). The topological structure of Figure S1c is recorded in node 4. In the following step,  $s_3$  is aligned to each segment of node 4. Again, the optimal group of consistent segments is picked out, and a topological structure depicting relationships of  $s_1$ ,  $s_2$  and  $s_3$  is created. The newly-created topological structure is recorded in node 5 and is shown in Figure S1d. By using the topological structure, the alternative splicing patterns can be illustrated in a simple way, which facilitates the use of graph theories to analyze the alternative splicing. From Figure 1d, the raw alignment of  $s_1$ ,  $s_2$  and  $s_3$  is obtained. With the location information of exons, which is annotated in GenBank, the alignment is then modified. This step makes the result precise around splice sites and avoids the loss of known information.

A paper describing the details in constructing the guide-tree and aligning sequences has been accepted for publication [Research on multiple alignment for alternative splicing. *Journal of Tsinghua University (Science and Technology)*, in press]. In the paper, the algorithm is analyzed to show how and why the performance is improved. Alignments of seven mRNAs of EP3 gene created by CLUSTALW, DIALIGN and Asalign are shown at <http://166.111.30.65/AsalignResult1.html> for comparison.

The alignment results are recorded in plain text files. Each result consists of four parts as described below:

1. The first part is indices of sequences. In this part, sequences are indexed in a format as 'seq1>AsHs21mRNA1|gnl|UGiHs#S625 Human activating transcription factor 3 (ATF3) mRNA, complete cds /cds=(164,709) /gb=L19871 /gi=442421 /ug=Hs.460 /len=1914'. The 'seqN' before '>' means the present

sequence is the  $N$ th sequence in the alignment. The text after '>' is the description of the sequence. This description includes the AsMamDB accession number of the cluster to which the sequence belongs, the molecular type of the sequence, the title of the sequence, the CDS in the sequence, the GenBank accession number, the GI identifier, the UniGene identifier and the length of the sequence.

2. The second part is the overview of the alignment. Figure S2 shows an example. The overview is divided into two parts by '[[[... >>>]]]'. The upper part is composed of mRNAs, while the lower part is composed of DNAs. The '[[[... >>>]]]' is used for indexing segments, which will be described in detail in Topological structure. In the overview, '\*\* P1\*( P2)\* P3\*\*' is used to illustrate the segment created by Asalign. P2 is the index of the sequence, to which the segment belongs. P1 and P3 are the start point and the end point of the segment in the sequence P2, respectively. If '\*\* P11\*( P12)\* P13\*\*', ... and '\*\* Pn1\*( Pn2)\* Pn3\*\*' overlap, the subsequence (P11, P13) of P12, ... and the subsequence (Pn1, Pn3) of Pn2 are common sequences.
3. The third part is the detail of alignment. A gap is designated by '-'.
4. The fourth part records some parameters calculated by the Asalign. These parameters include the total length, average length, segment number and the average ratio of the common parts of the sequences.

Each alignment is linked to a topological structure, which is displayed by a Java applet. The topological structure can be accessed by clicking the uppermost hyperlink 'Topological Structure'.

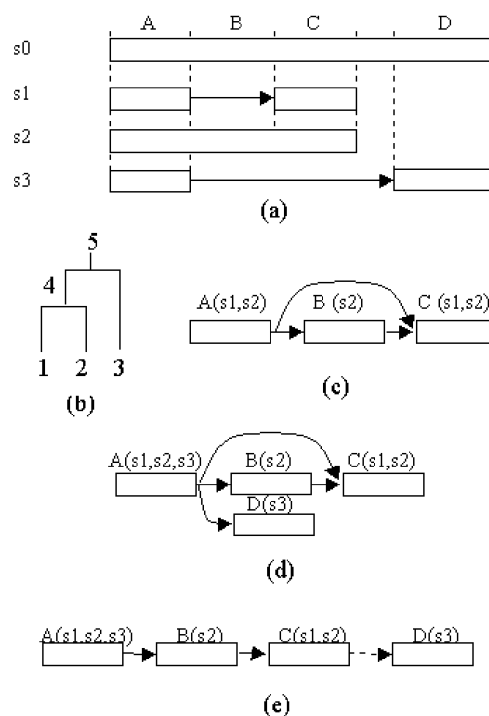


Figure S1. The principles of Asalign.

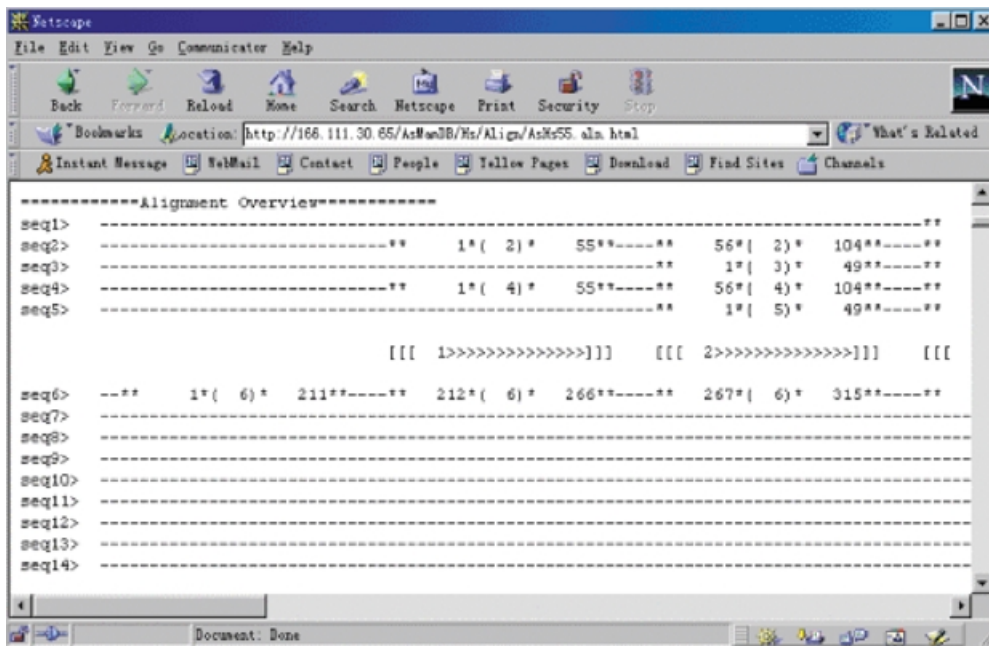


Figure S2. The overview of alignment of mRNAs and DNAs of CTSE gene of human.

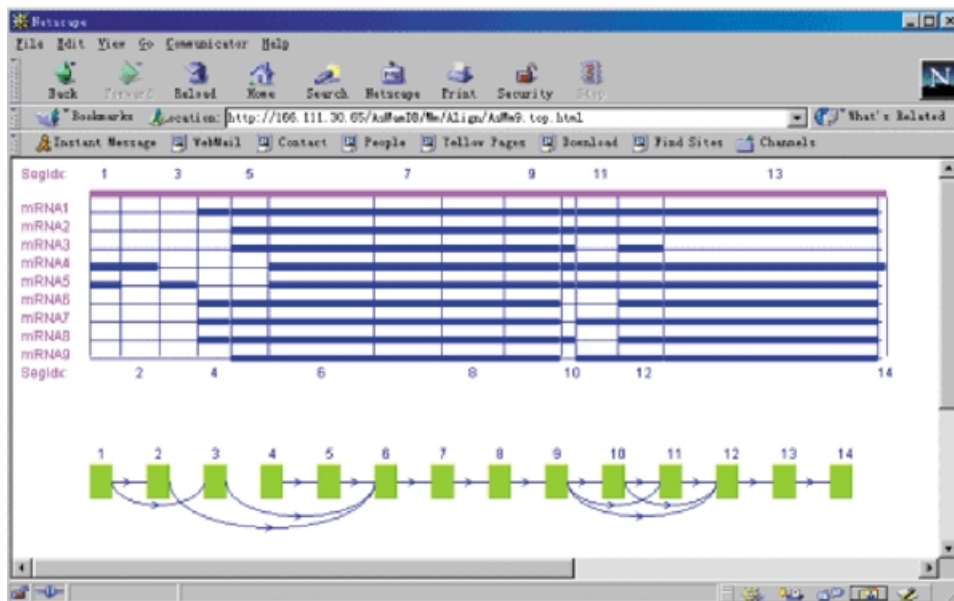
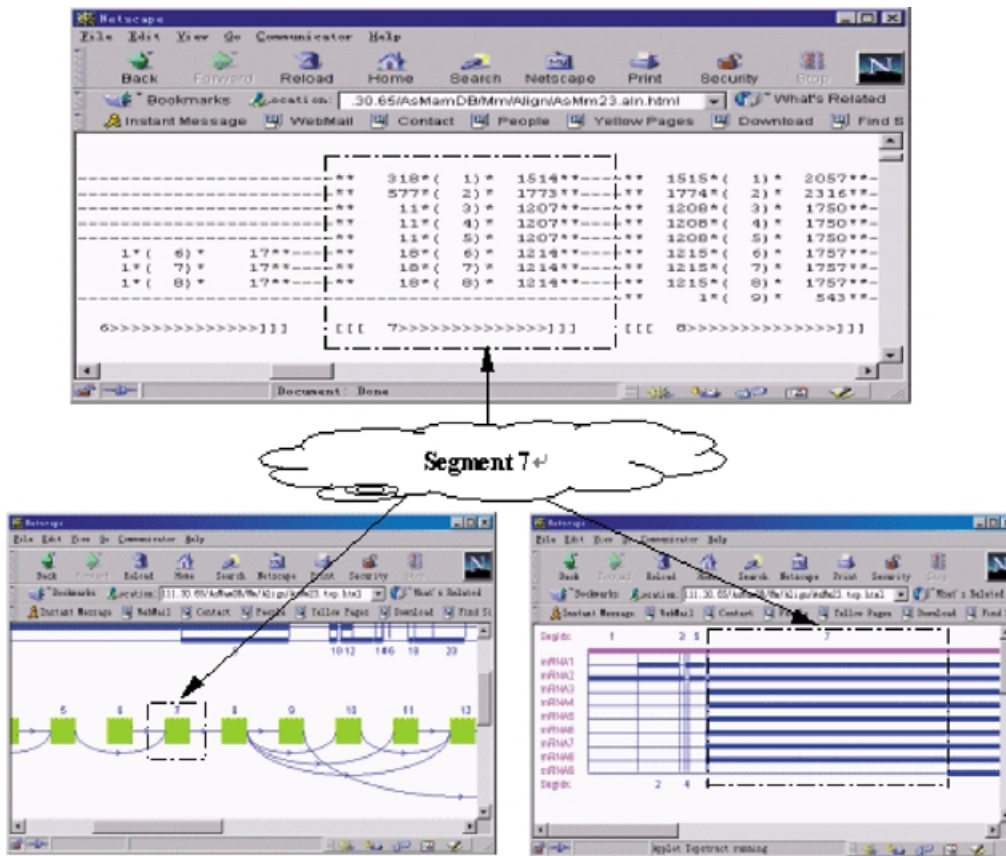


Figure S3. The topological structure of Mef2b gene of mouse.

### Topological structure

The topological structure is mainly used for analysis using graph theory, which may be of help in the study of alternative splicing patterns. The alignment created by Asalign is usually composed of a set of gap-free segments, and the relationships of these segments can be shown in a topological structure. The procedure of creating the topological structure is described in the first section.

Using the topological structure, the types of alternatively spliced exons can be illustrated. Taking Figure S3 as an example, the eleventh segment is skipped by some sequences. This segment is a skipped exon, if it is a complete exon. If this segment is the first one of an exon, which is composed of several segments, it is an alternative start of an exon. If this segment is the last one of an exon, which is composed of several segments, it is an alternative start of an intron. The relationship between segments



**Figure S4.** An example of segment annotation in topological structure graph. The segments indicated in rectangles are the same segment in alignment and topological structure.

and exons can be obtained from gene structure and alignment files provided by the database.

In order to make the topological structure more similar to the gene's real structure, we have refined it as follows. Assuming that segments A and B come from the same group of sequences, and A is previous to B, the end points of A are  $(e_1, e_2, \dots, e_m)$ , and the start points of B are  $(h_1, h_2, \dots, h_m)$ . Let  $d = \max\{|h_1 - e_1|, h_2 - e_2|, \dots, h_m - e_m|\}$ . If  $d$  is no larger than a threshold, A and B will be regarded as one segment. This refinement avoids breaking one exon into two or more segments if insertions or deletions of bases introduce small gaps into the alignment. Refined segments are labeled in the alignment and in the topological structure as shown in Figure S4.

The topological structure consists of the following three parts:

1. The first part is an alignment overview of mRNAs. Unlike the overview provided by alignment file, this overview does not contain DNAs, and is displayed by Java applet. The first displayed sequence is a template sequence, which is composed of all the segments created by Asalign.
2. The second part is the topological structure, created by Asalign and refined by the method mentioned above. In the topological structure, an arrow from one segment to another means the first segment comes before the second.
3. The third part is the indices of the sequences. The contents are the same as the first part of the alignment file.

Each topological structure is linked to a segment file, which records the detailed information of segments. In this file, nucleotide sequences of each segment are provided. The file also reports indices of mRNAs and DNAs contained in a specific segment. For example, '>segment1 | 1,3,' means the first segment exists in the first and third sequences of the cluster.

It must be pointed out that the topological structure currently used may break one exon into several segments. This is not convenient for analyzing the types of exons. We are designing a better topological structure. The refined topological structure will display the types of exons directly, and will replace the old one as soon as it is available.

## GENE STRUCTURE

The gene structure is illustrated by the distribution of exons, introns, promoters, enhancers and other regulatory elements in the alignment overview. Location information of these elements is retrieved from GenBank. The start points are compared segment to segment with the alignment overview. If the start point of an element is located in the range defined by the start and end points of a segment, the element is located in this segment.

The gene structure files provided for users are also plain text files. Each file consists of two parts as described below:

1. The first part is indices of sequences. The contents are the same as the first part of alignment.
2. The second part is the gene structure. The left part of the gene structure is the alignment overview created by Asalign. It is displayed from upside to downside, corresponding to the alignment overview from left to right. Above the alignment overview, there are indices of sequences. From left to right, the indices are 1, 2, 3 and so on. In order to indicate the start and the end points of each segment, symbols like '====>s1-(n1) s2-(n2) s3-(n3)' are used. For example, the symbol '====>2-(55) 4-(55) 6-(266)' means the point indicated by the arrow is the 55th base of the second sequence, the 55th base of the fourth sequence and the 266th base of the sixth sequence. The right part of the gene structure is the distribution of exons, introns, promoters, enhancers and other regulatory elements.

Each functional element is shown as '<<[...](+/-)'. For example, '<<[5'UTR=AJ250716.1:1..104](+)' means, the position indicated by '<<' is the start point of a 5'UTR, which is annotated in GenBank entry AJ250716 and stretches from the first base to the 104th base of this entry. If the element is located in the complementary sequence of the alignment overview, it is labeled with (-), otherwise with (+).

## REFERENCES

1. Thompson, J.D., Higgins D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
2. Morgenstern, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, **15**, 211–218.

# Asalign-comparison with CLUSTALW, DIALIGN(1)

To Compare the Asalign with CLUSTALW and DIALIGN, alignments of seven mRNAs of EP3 gene are given below.  
The seven mRNAs are:

```
mRNA1:join(D86087.1:361..1488,D86088.1:38..3497)
mRNA2:join(D86087.1:361..1488,D86088.1:38..304,D86091.1:26..52,D86096.1:34..451)
mRNA3:join(D86087.1:361..1488,D86088.1:38..304,D86091.1:26..52,D86094.1:35..179,D86095.1:20..112,D86096.1:34..451)
mRNA4:join(D86087.1:361..1488,D86088.1:38..304,D86093.1:31..106,D86095.1:20..112,D86096.1:34..451)
mRNA5:join(D86087.1:361..1488,D86088.1:38..304,D86091.1:26..52,D86093.1:31..106,D86095.1:20..112,D86096.1:34..451)
mRNA6:join(D86087.1:361..1488,D86088.1:38..304,D86096.1:34..451)
mRNA7:join(D86087.1:361..1488,D86088.1:38..304,D86089.1:33..124,D86091.1:26..52,D86096.1:34..451)
```

## 1. Alignment created by CLUSTALW

CLUSTAL W (1.8) multiple sequence alignment

```
mRNA4      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
mRNA5      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
mRNA3      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
mRNA2      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
mRNA7      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
mRNA6      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
mRNA1      ACCAGAGGTTTCCAGAGAGGAAGGCGTGGCTCCCTCCCGGGCCAGTGAGCCCTGGCGCC
*****
```

```
mRNA4      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
mRNA5      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
mRNA3      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
mRNA2      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
mRNA7      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
mRNA6      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
mRNA1      GCCGCGGCCGCGGTCCCAGCAGCGGAGTAGGGCGGCGGCTGCGCCCCGCACCATGGGGGG
*****
```

```
mRNA4      CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCGCCCGCGCCGCGTCTGCCC
mRNA5      CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCGCCCGCGCCGCGTCTGCCC
mRNA3      CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCGCCCGCGCCGCGTCTGCCC
mRNA2      CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCGCCCGCGCCGCGTCTGCCC
mRNA7      CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCGCCCGCGCCGCGTCTGCCC
```

mRNA6 CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCCGCCGCGCCGCTCTGCCC  
mRNA1 CAGCCCAGCCCCAGCCGCGGTAAACGCCGACCTCCGCCGCCCGCCGCGCCGCTCTGCCC  
\*\*\*\*\*

mRNA4 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
mRNA5 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
mRNA3 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
mRNA2 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
mRNA7 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
mRNA6 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
mRNA1 CCTCCCGCTGCGGCTCTCTGGACGCCATCCCCCTCTCACCTCGAAGCCAACATGAAGGAG  
\*\*\*\*\*

mRNA4 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
mRNA5 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
mRNA3 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
mRNA2 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
mRNA7 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
mRNA6 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
mRNA1 ACCCGGGGCTACGGAGGGGATGCCCCCTTCTGCACCCGCCTCAACCACTCCTACACAGGC  
\*\*\*\*\*

mRNA4 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
mRNA5 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
mRNA3 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
mRNA2 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
mRNA7 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
mRNA6 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
mRNA1 ATGTGGGCGCCCGAGCGTTCCGCCGAGGCGCGGGGCAACCTCACGCGCCCTCCAGGGTCT  
\*\*\*\*\*

mRNA4 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
mRNA5 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
mRNA3 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
mRNA2 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
mRNA7 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
mRNA6 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
mRNA1 GGCGAGGATTGCGGATCGGTGTCCGTGGCCTTCCCAGATCACCATGCTGCTCACTGGTTTC  
\*\*\*\*\*

mRNA4 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCGAGCTACCGGCGCCGGGAGAGCAAG  
mRNA5 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCGAGCTACCGGCGCCGGGAGAGCAAG  
mRNA3 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCGAGCTACCGGCGCCGGGAGAGCAAG  
mRNA2 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCGAGCTACCGGCGCCGGGAGAGCAAG  
mRNA7 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCGAGCTACCGGCGCCGGGAGAGCAAG  
mRNA6 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCGAGCTACCGGCGCCGGGAGAGCAAG

mRNA1 GTGGGCAACGCACTGGCCATGCTGCTCGTGTGCGCAGCTACCGGCGCCGGGAGAGCAAG  
\*\*\*\*\*

mRNA4 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
mRNA5 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
mRNA3 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
mRNA2 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
mRNA7 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
mRNA6 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
mRNA1 CGCAAGAAGTCCTTCCCTGCTGTGCATCGGCTGGCTGGCGCTACCGACCTGGTCGGGCAG  
\*\*\*\*\*

mRNA4 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
mRNA5 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
mRNA3 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
mRNA2 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
mRNA7 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
mRNA6 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
mRNA1 CTTCTCACCACCCGGTCGTCATCGTTCGTGTACCTGTCCAAGCAGCGTTGGGAGCACATC  
\*\*\*\*\*

mRNA4 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
mRNA5 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
mRNA3 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
mRNA2 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
mRNA7 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
mRNA6 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
mRNA1 GACCCGTCGGGGCGGCTCTGCACCTTTTTTCGGGCTGACCATGACTGTTTTTCGGGCTCTCC  
\*\*\*\*\*

mRNA4 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
mRNA5 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
mRNA3 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
mRNA2 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
mRNA7 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
mRNA6 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
mRNA1 TCGTTGTTTCATCGCCAGCGCCATGGCCGTCGAGCGGGCGCTGGCCATCAGGGCGCCGCAC  
\*\*\*\*\*

mRNA4 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG  
mRNA5 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG  
mRNA3 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG  
mRNA2 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG  
mRNA7 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG  
mRNA6 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG  
mRNA1 TGGTATGCGAGCCACATGAAGACGCGTGCCACCCGCGCTGTGCTGCTCGGCGTGTGGCTG

\*\*\*\*\*

mRNA4 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
mRNA5 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
mRNA3 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
mRNA2 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
mRNA7 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
mRNA6 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
mRNA1 GCCGTGCTCGCCTTCGCCCTGCTGCCGGTGCTGGGCGTGGGCCAGTACACCGTCCAGTGG  
\*\*\*\*\*

mRNA4 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
mRNA5 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
mRNA3 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
mRNA2 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
mRNA7 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
mRNA6 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
mRNA1 CCCGGGACGTGGTGCTTCATCAGCACCGGGCGAGGGGGCAACGGGACTAGCTCTTCGCAT  
\*\*\*\*\*

mRNA4 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
mRNA5 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
mRNA3 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
mRNA2 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
mRNA7 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
mRNA6 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
mRNA1 AACTGGGGCAACCTTTTTCTTCGCCTCTGCCTTTGCCTTCCTGGGGCTCTTGGCGCTGACA  
\*\*\*\*\*

mRNA4 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
mRNA5 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
mRNA3 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
mRNA2 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
mRNA7 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
mRNA6 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
mRNA1 GTCACCTTTTTCCTGCAACCTGGCCACCATTAAGGCCCTGGTGTCCCCTGCGGGCCAAG  
\*\*\*\*\*

mRNA4 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
mRNA5 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
mRNA3 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
mRNA2 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
mRNA7 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
mRNA6 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
mRNA1 GCCACGGCATCTCAGTCCAGTGCCAGTGGGGCCGCATCACGACCGAGACGGCCATTCAG  
\*\*\*\*\*



mRNA4 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
mRNA5 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
mRNA3 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
mRNA2 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
mRNA7 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
mRNA6 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
mRNA1 CTTATGGGGATCATGTGCGTGTGTCGGTCTGCTGGTCTCCGCTCCTGATAATGATGTTG  
\*\*\*\*\*

mRNA4 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
mRNA5 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
mRNA3 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
mRNA2 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
mRNA7 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
mRNA6 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
mRNA1 AAAATGATCTTCAATCAGACATCAGTTGAGCACTGCAAGACACACACGGAGAAGCAGAAA  
\*\*\*\*\*

mRNA4 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
mRNA5 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
mRNA3 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
mRNA2 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
mRNA7 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
mRNA6 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
mRNA1 GAATGCAACTTCTTCTTAATAGCTGTTTCGCCTGGCTTCACTGAACCAGATCTTGGATCCT  
\*\*\*\*\*

mRNA4 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
mRNA5 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
mRNA3 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
mRNA2 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
mRNA7 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
mRNA6 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
mRNA1 TGGGTTTACCTGCTGTTAAGAAAGATCCTTCTTCGAAAGTTTGGCCAGGTAGCAAATGCT  
\*\*\*\*\*

mRNA4 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
mRNA5 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
mRNA3 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
mRNA2 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
mRNA7 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
mRNA6 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
mRNA1 GTCTCCAGCTGCTCTAATGATGGACAGAAAGGGCAGCCTATCTCATTATCTAATGAAATA  
\*\*\*\*\*

mRNA4 ATACAGACAGAAGCA-----  
mRNA5 ATACAGACAGAAGCAATGAGAAAAAGAA--GACTCAGAGAGCAA-----  
mRNA3 ATACAGACAGAAGCAATGAGAAAAAGAA--GACTCAGAGAGCAAGCTCCTCTTCTTCCCA  
mRNA2 ATACAGACAGAAGCA-----  
mRNA7 ATACAGACAGAAGCA-----  
mRNA6 ATACAGACAGAAGCA-----  
mRNA1 ATACAGACAGAAGCATGAAAGAAAACACTTAACTTGCATGTGCACAGCTTTTGGTAACAA  
\*\*\*\*\*

mRNA4 -----GAGATGGGGCCTGA--TGGAAGGTGT-----TTTTGTCA-TGCAT---  
mRNA5 -----GAGATGGGGCCTGA--TGGAAGGTGT-----TTTTGTCA-TGCAT---  
mRNA3 CCCCTACTGTGATTGATCCTTC--AAGGTTCTGTGCTCAGCCCTTCCGTTGGTTCTT---  
mRNA2 -----  
mRNA7 -----ATCAGGTACCA--CACAA-----A-----CAACTA-TGCAT---  
mRNA6 -----  
mRNA1 ATATCGCTAAACCTTACTGTGAATTTAGGCATCTCTGGCATGCCACTGTTTATGCATTGA

mRNA4 ----GGA-----GGCA--GGTCC-----CCAGGACT---TGGTGCA  
mRNA5 ----GGA-----GGCA--GGTCC-----CCAGGACT---TGGTGCA  
mRNA3 ----GGATTTGTCCTTTCCCGCCA--TGTCTTCATCACATCCACAACACT---TCCACTA  
mRNA2 -----A-----  
mRNA7 ----CCA-----GCTC--CACCT-----CCTTACCC---TG--CCA  
mRNA6 -----  
mRNA1 AGTGGAATTTTTGGTATAAAGCTAAATGGTCTTAGAAGCATAGAAAATCCCTATGTGCCA

mRNA4 GTTCTCATGA-----TAGAGAACCCTG-----CAGTGTCCA-GCTAAG  
mRNA5 GTTCTCATGA-----TAGAGAACCCTG-----CAGTGTCCA-GCTAAG  
mRNA3 ACACTTGCAGCTTCAAACCTTCTTAGAGAACCCTG-----CAGTGTCCA-GCTAAG  
mRNA2 -----TGA-----  
mRNA7 GTGTTCTCA-----ACCTTG-----ATGTG-----GAG  
mRNA6 -----  
mRNA1 AAAGTAGTGAAACACAAACAAAGGAAAATATATTAATAACAGTCTAGTGTTTTTGTGAG

mRNA4 CTGATGACTTGAAGATAAAATCTGC--CTAACCCCTGGGATGAAG----TATCTGTGAACTA  
mRNA5 CTGATGACTTGAAGATAAAATCTGC--CTAACCCCTGGGATGAAG----TATCTGTGAACTA  
mRNA3 CTGATGACTTGAAGATAAAATCTGC--CTAACCCCTGGGATGAAG----TATCTGTGAACTA  
mRNA2 -----GAAAAAGAAG-----A  
mRNA7 C-GACCATTTGAAA-GA-----TGAGAAAAAG----AAG-----A  
mRNA6 -----G-----  
mRNA1 TCTGCCATTCGTAGCTGAATATGTGATTAATTATGTGATGAAAACATTTTTTATAAATGA

mRNA4 TTTTGAC----AGCAGATGAGGA-----ATTTTGGGGAAATTAAACCTGCCTT-T

mRNA5 TTTTGAC-----AGCAGATGAGGA-----ATTTTGGGGAAATTTAAACCTGCCTT-T  
 mRNA3 TTTTGAC-----AGCAGATGAGGA-----ATTTTGGGGAAATTTAAACCTGCCTT-T  
 mRNA2 CTCAGAG-----AGCAA--GAGGA-----ATTTTGGGGAAATTTAAACCTGCCTT-T  
 mRNA7 CTCAGAG-----AGCAA--GAGGA-----ATTTTGGGGAAATTTAAACCTGCCTT-T  
 mRNA6 -----A---GGA-----ATTTTGGGGAAATTTAAACCTGCCTT-T  
 mRNA1 TCTTGGTCTATTGGGGAGCGGGGATAGTTAATATTCCAGTACACTGAATACATGAGGAAT  
 \*\*\* \*\* \* \* \* \* \* \*

mRNA4 CTGCCAGGATCACATCACTGGAAGCTCC--ATGACTCTCTTTTTGTAAAAGAAAAAAAAA  
 mRNA5 CTGCCAGGATCACATCACTGGAAGCTCC--ATGACTCTCTTTTTGTAAAAGAAAAAAAAA  
 mRNA3 CTGCCAGGATCACATCACTGGAAGCTCC--ATGACTCTCTTTTTGTAAAAGAAAAAAAAA  
 mRNA2 CTGCCAGGATCACATCACTGGAAGCTCC--ATGACTCTCTTTTTGTAAAAGAAAAAAAAA  
 mRNA7 CTGCCAGGATCACATCACTGGAAGCTCC--ATGACTCTCTTTTTGTAAAAGAAAAAAAAA  
 mRNA6 CTGCCAGGATCACATCACTGGAAGCTCC--ATGACTCTCTTTTTGTAAAAGAAAAAAAAA  
 mRNA1 TTAACCACAT-ACATCATTGAAGACAAGGGATAGCAGTTTGTTTTTATTCAAAGACATTG  
 \* \* \*\* \*\*\*\*\* \* \* \* \*\* \* \* \* \* \* \* \* \* \* \*

mRNA4 TCACAGAAACACCCACCTCCCAAATATTCTCTTTTACTTCTTCCCCCA--AGCCACCC  
 mRNA5 TCACAGAAACACCCACCTCCCAAATATTCTCTTTTACTTCTTCCCCCA--AGCCACCC  
 mRNA3 TCACAGAAACACCCACCTCCCAAATATTCTCTTTTACTTCTTCCCCCA--AGCCACCC  
 mRNA2 TCACAGAAACACCCACCTCCCAAATATTCTCTTTTACTTCTTCCCCCA--AGCCACCC  
 mRNA7 TCACAGAAACACCCACCTCCCAAATATTCTCTTTTACTTCTTCCCCCA--AGCCACCC  
 mRNA6 TCACAGAAACACCCACCTCCCAAATATTCTCTTTTACTTCTTCCCCCA--AGCCACCC  
 mRNA1 CTGTGTTCTCTTTTCATTGCCTCTCTCGCTTTCTGTCACTTTTTTCCTCCTTACATTAAG  
 \* \*\* \*\* \* \*\*\* \* \*\*\*\*\* \* \* \* \* \* \* \* \*

mRNA4 CCAAATATAACTGTTATCCAGAAGCTGTTATGTCCTGTTTCCATACAT-GTTTTTGTACT  
 mRNA5 CCAAATATAACTGTTATCCAGAAGCTGTTATGTCCTGTTTCCATACAT-GTTTTTGTACT  
 mRNA3 CCAAATATAACTGTTATCCAGAAGCTGTTATGTCCTGTTTCCATACAT-GTTTTTGTACT  
 mRNA2 CCAAATATAACTGTTATCCAGAAGCTGTTATGTCCTGTTTCCATACAT-GTTTTTGTACT  
 mRNA7 CCAAATATAACTGTTATCCAGAAGCTGTTATGTCCTGTTTCCATACAT-GTTTTTGTACT  
 mRNA6 CCAAATATAACTGTTATCCAGAAGCTGTTATGTCCTGTTTCCATACAT-GTTTTTGTACT  
 mRNA1 AAAAGTTAATTACAGTTAAAAATGTATAATGTATTTATAATTCATCGATACCATTAT  
 \*\* \*

mRNA4 TTTACTATATCTACATACATCAATTAAACTTATGTCCTATTG-----TTTTGTGA  
 mRNA5 TTTACTATATCTACATACATCAATTAAACTTATGTCCTATTG-----TTTTGTGA  
 mRNA3 TTTACTATATCTACATACATCAATTAAACTTATGTCCTATTG-----TTTTGTGA  
 mRNA2 TTTACTATATCTACATACATCAATTAAACTTATGTCCTATTG-----TTTTGTGA  
 mRNA7 TTTACTATATCTACATACATCAATTAAACTTATGTCCTATTG-----TTTTGTGA  
 mRNA6 TTTACTATATCTACATACATCAATTAAACTTATGTCCTATTG-----TTTTGTGA  
 mRNA1 TCAAATATTGCTCAATACAGCAAATTAGCTCCTAACCTAACAAAGTTTAAGTTTACTTGG  
 \*

mRNA4 ATTTATATTTGCGTATACATTAT-CATATGTAAAATTTGCATTTTTTTTATTGAAAATTAT  
 mRNA5 ATTTATATTTGCGTATACATTAT-CATATGTAAAATTTGCATTTTTTTTATTGAAAATTAT

mRNA3 ATTTATATTTGCGTATACATTAT-CATATGTAAAATTTGCATTTTTTTTATTGAAAATTAT  
mRNA2 ATTTATATTTGCGTATACATTAT-CATATGTAAAATTTGCATTTTTTTTATTGAAAATTAT  
mRNA7 ATTTATATTTGCGTATACATTAT-CATATGTAAAATTTGCATTTTTTTTATTGAAAATTAT  
mRNA6 ATTTATATTTGCGTATACATTAT-CATATGTAAAATTTGCATTTTTTTTATTGAAAATTAT  
mRNA1 ATTGATAATTAGGTTTACTCTTTATCTGAATAAGAACCAATTCATTTGTTTGAAAT-AT  
\*\*\* \*\* \* \* \*\* \*

mRNA4 GTTTCTTGAGATTTATCCACATTGAAACATGGAGCTC-----TAAATCGTTAATTTT  
mRNA5 GTTTCTTGAGATTTATCCACATTGAAACATGGAGCTC-----TAAATCGTTAATTTT  
mRNA3 GTTTCTTGAGATTTATCCACATTGAAACATGGAGCTC-----TAAATCGTTAATTTT  
mRNA2 GTTTCTTGAGATTTATCCACATTGAAACATGGAGCTC-----TAAATCGTTAATTTT  
mRNA7 GTTTCTTGAGATTTATCCACATTGAAACATGGAGCTC-----TAAATCGTTAATTTT  
mRNA6 GTTTCTTGAGATTTATCCACATTGAAACATGGAGCTC-----TAAATCGTTAATTTT  
mRNA1 GGAGTTTGTGACT-ACCCAAATTGCTAATTATCTTTCTTTTGAATATATTTTACATTTT  
\* \*\* \*

mRNA4 AACCG--CTATAGAGTATTCATAATTTGAATAA--AGCATAATTTGTTTGTAC-----  
mRNA5 AACCG--CTATAGAGTATTCATAATTTGAATAA--AGCATAATTTGTTTGTAC-----  
mRNA3 AACCG--CTATAGAGTATTCATAATTTGAATAA--AGCATAATTTGTTTGTAC-----  
mRNA2 AACCG--CTATAGAGTATTCATAATTTGAATAA--AGCATAATTTGTTTGTAC-----  
mRNA7 AACCG--CTATAGAGTATTCATAATTTGAATAA--AGCATAATTTGTTTGTAC-----  
mRNA6 AACCG--CTATAGAGTATTCATAATTTGAATAA--AGCATAATTTGTTTGTAC-----  
mRNA1 TATGAGCCTAAGGAAGATTCTAGAACTGACCTATGAGAGTCGTGAAGTCGTTTTTCAGA  
\* \*\* \*

mRNA4 -----  
mRNA5 -----  
mRNA3 -----  
mRNA2 -----  
mRNA7 -----  
mRNA6 -----  
mRNA1 ATGCTATGTAAGGACCGATTTGAGCACTAACTATAGGTA CTCTGAATATATATTTCCCTT

mRNA4 -----  
mRNA5 -----  
mRNA3 -----  
mRNA2 -----  
mRNA7 -----  
mRNA6 -----  
mRNA1 GATTATTCACCAAAGTGTTCCCCAGTCTTTGACTCTTTAAATTCCAATACTGATTCCAA

mRNA4 -----  
mRNA5 -----  
mRNA3 -----