SUPPLEMENTARY DATA

CONBIND ALGORITHM

The pseudo code of the MSA algorithm described in the main text is shown below.

Algorithm 1 *Given a list of sequences (sequencesList) and a list of motifs families (motifsList), the following algorithm produces an aligned list of sequences using the motifs information to optimize the alignment.*

```
symbolsList = emptyList()
ForEach sequence in sequencesList
    ForEach motif in motifsList
        subsequencesList = searchForMotifs(sequence, getPatterns(motif))
        ForEach subsequence in subsequencesList
            ForEach base in subsequence
                symbol = getSymbolForBase(base, getFamily(motif))
                symbolsList.add(symbol)
                replaceLetter(subsequence, base, symbol)
            EndFor
        EndFor
    EndFor
EndFor
extendedMatrix = getExtendedMatrix(symbolsList)
alignedSequences = performMSA(sequencesList, extendedMatrix)
ForEach sequence in alignedSequences
    ForEach symbol in symbolsList
        base = getBaseForSymbol(symbol)
        replaceLetter(sequence, symbol, base)
    EndFor
EndFor
Return alignedSequences
```

WEIGHTS ESTIMATION

The performance of our motif-aware alignment method depends on two parameters; the motif match weight (MMW) and the motif mismatch weight (MSW), as described in the method section of the main text. These two parameters are responsible for the trade-off between overall alignment score and the number of aligned TFBSs. Intuitively, the heavier the MMW and the MSW, the more TFBSs will be aligned on the expense of the overall alignment score. Furthermore, in case of extremely heavy weights, TFBSs that lay far apart on their respective

regulatory regions, can be forcefully pulled together in the alignment, artificially aligning unrelated TFBSs. Contrarily, the weaker the MMW and the MSW are, the more likely it will be for the algorithm to discard the information about motifs and optimize the alignments in the traditional way (i.e. maximising the alignment score). An optimal motif-aware alignment method should produce alignments with a minimal change in alignment score and, at the same time, be able to align all (and only) functional TFBSs. These two objectives, however, are clearly discordant. We can imagine the difference in alignment score as the Cost that we need to pay for a certain gain in Effectiveness, i.e. the number of functional TFBSs correctly aligned.

In order to find the best parameter values and assessing the quality of the produced alignment compared with the efficiency in identifying conserved TFBSs we trained our method using a set of regulatory regions for which the functional TFBSs where previously experimentally validated by the Göttgens Lab in murine cell lines (personal communication). This training set includes 14 regulatory regions (Erg+75, Erg+65, Scl+40, Cx3cr1 promoter, Gfi1b+16 Meis1+48, Gfi1b+17, Pim1+10, Lmo2-70, Scl+19, Lyl1+2, Fli1-15, Gata2-3, PU.1-14) for a total of 114 experimentally validated TFBSs belonging to six motif families (i.e. ETS, GATA, EBOX, GFI1, MEIS, and RUNT). These 14 mouse regions were aligned to seven different organisms (i.e. *Homo sapiens*, *Bos taurus*, *Canis lupus familiaris*, *Loxodonta africana*, *Monodelphis domestica*, *Sarcophilus harrisii*, and *Ornithorhynchus anatinus*) using ConBind. We performed an exhaustive parameter sweep running ConBind with different MMW and MSW pairs, such that $1 \leq MMW \leq 50$ and $0 \leq MSW \leq MMW$ (notice that assigning a heavier weight to a mismatch would not be a sensible option), for a total of 1325 runs. The sum-of-pairs score as defined by Thompson *et al.* (1999) was used to assess the overall quality of the produced alignments. For each ConBind run (with a specific MMW and MSW pair) two values were computed for each region *R*. The cost

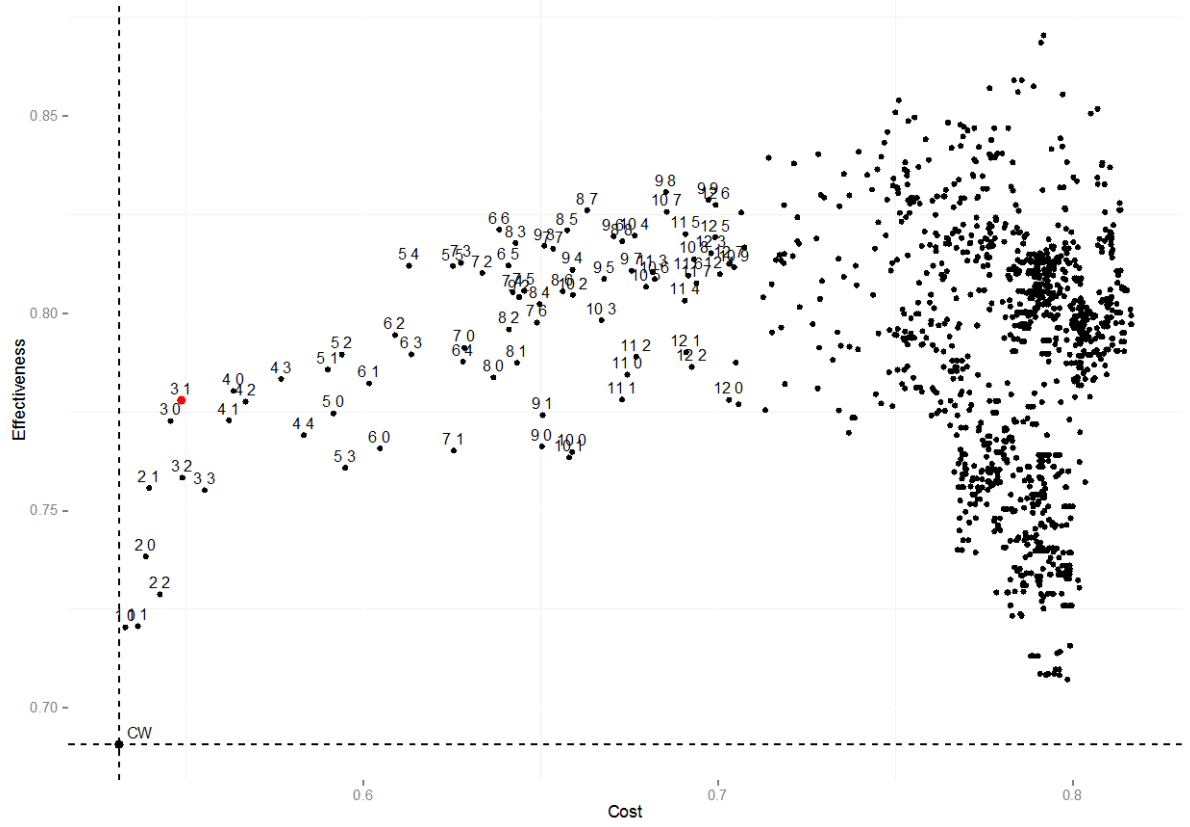
$$Cost_R\left(A_R^{MMW,MSW}\right) = 1 - Sum - of - pairs\ Score(A_R^{MMW,MSW}),$$

where $A_R^{MMW,MSW}$ is the alignment of the region *R* produced by ConBind using the weight pair MMW,MSW. The columns corresponding to TFBSs were excluded in the Cost calculation. Notice that a perfect alignment yields a sum-of-pairs score of 1. For a perfect alignment *A* we expect $Cost_R\left(A_R^{MMW,MSW}\right)$ to tend to 0. The second value we computed is the effectiveness

$$\text{Effectiveness}_R\left(A_R^{MMW,MSW}\right) = \frac{\sum_{T_R^i}^{T_R} \frac{n(T_R^i)}{H}}{T_R}$$

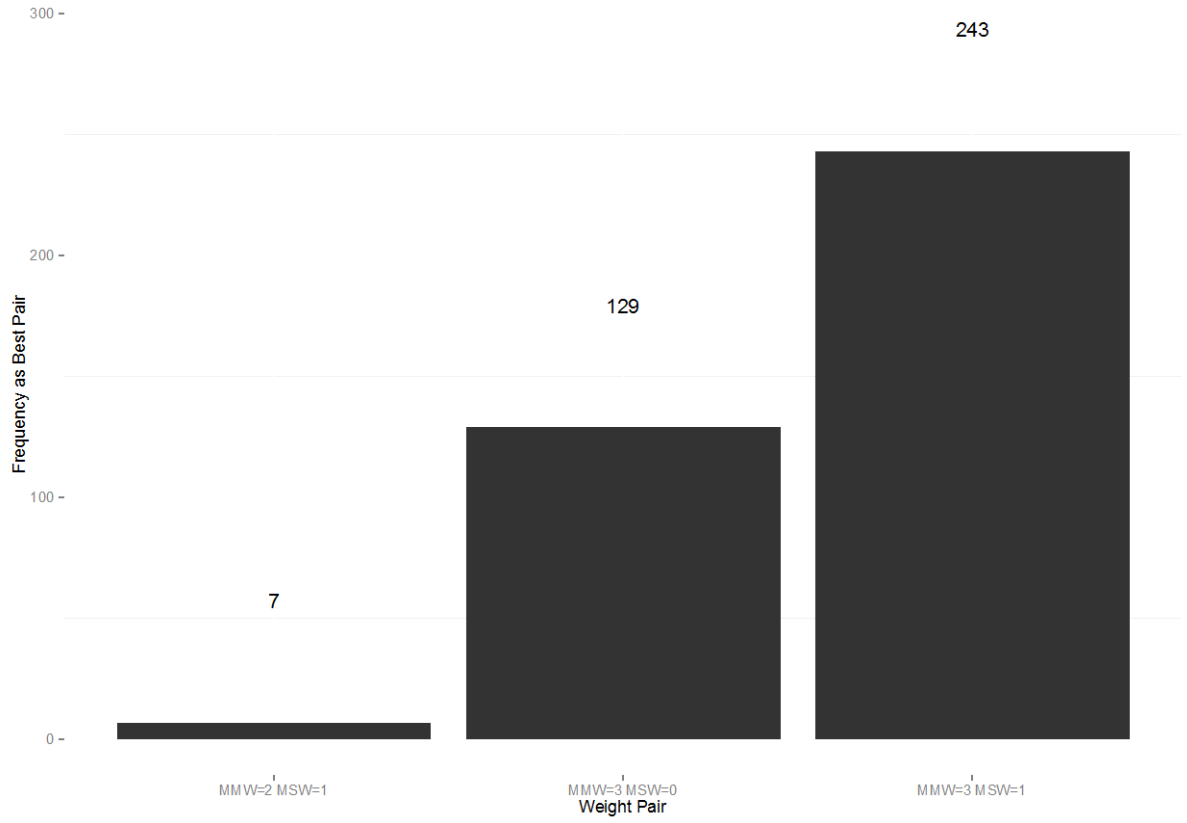
where T_R is the total number of experimentally validated TFBS on the mouse region *R*. T_R^i is the i^{th} experimentally validated TFBS on the mouse sequence, *H* the number of sequences used in the alignment and $n(T_R^i)$ is the number of organism (including mouse) in which T_R^i is aligned in the same position. Notice that for every i, $n(T_R^i)/H$ should approach 1, since all T_R have been experimentally validated and should be conserved in *H* sequences. Therefore, we expect $\text{Effectiveness}_R\left(A_R^{MMW,MSW}\right)$ to tend to 1 for an alignment A that shows highly conserved T_R. Supplementary Figure 1 shows the mean Cost and Effectiveness (over all 14 regions) computed

for every MMW, MSW combination. Notably, every weight combination shows an improvement in terms of efficiency over ClustalW2 alignments of the same regions.



Supplementary Figure 1 *Mean cost and effectiveness of 14 benchmark regions for every MMW, MSW combination. Each dot corresponds to a weight pair, where the number on the left is the MMW and the MSW is on the right. The point labelled CW shows the cost and effectiveness computed using ClustalW2. The dot in red shows the weight pair with the best cost-effectiveness ratio after cross-validation.*

To reduce overtraining we computed cost and effectiveness using different subsets of the 14 regions. Specifically, we used every possible subset using 100%, 90% and 80% of the 14 regions, for a total of 379 subsets. For each subset we chose the MMW-MSW pair with the best cost-effectiveness ratio. For roughly 64% of the subsets the best weight pair had a MMW=3 and MSW=1. Supplementary Figure 2 shows the best weight pairs for the entire cross-validation study.

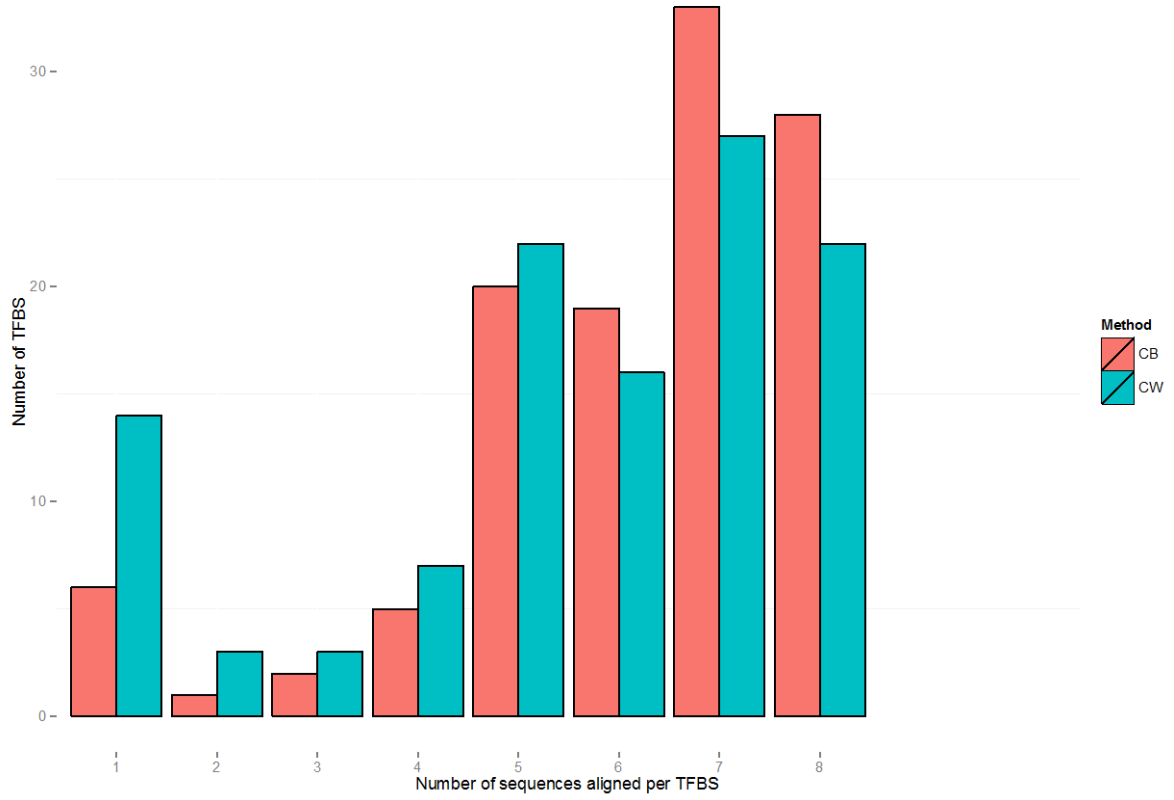


Supplementary Figure 2 *Only three weight pairs (over the total of 1325 tested) attained the best cost-efficiency ratio in at least one of the cross-validation subsets. Particularly, the pair with MMW=3 and MSW=1 is the one with the best cost-efficiency ratio for roughly 64% of the cross-validation subsets.*

Using the weight pair selected by cross-validation (i.e. MMW=3 and MSW=1) we compared the alignment of experimentally validated TFBSs between ConBind and ClustalW2, as shown in Supplementary Figure 3.

PRALINE AND PROGRESSIVE MULTIPLE SEQUENCE ALIGNMENT

The progressive multiple alignment step was performed with a flexible sequence alignment program, a reimplementation of the available PRALINE MSA toolbox (Heringa, 1999). This tool was developed in-house and thus has support for required features such as the use of custom symbol alphabets and weight matrices during the alignment process. In order to reduce the number of parameters, we used default settings and implemented a minimal tree-guided progressive alignment strategy. A reasonable default was chosen for the linkage method during the hierarchical clustering (UPGMA).



Supplementary Figure 3 *For each experimentally validated TFBS we counted the number of species in which the TFBS was aligned in the same position. ConBind (CB) can detect a higher conservation signal compared to ClustalW2 (CW), indicating that ConBind aligns more TFBSs.*

GENERATING THE GFI1B+13 CONSTRUCTS

Gfi1b+13_wt (chr2:28,457,606-28,458,256; mm9)

CAGGTGCTAGATCCCGTCATTTGGGACCAATACCTAGTTGTCCCTAGTAAATTTATGTCTACAGGGACCTGGAAC
CTTGGCAGTTAGAACAGAATTTCCTAGGTAGAGCAGGGCCCTGCCTTAGGAACTGAGATCTGGACAGTGGACACT
TGACTCTTCCTAGGACACACAGAATTAGTTCTGGGAAGATGCCACCCCAGTGGCCCCCATAGATCTAGCTGGGGT
TGAGCCCTGCCAGGAGCCAGGCTGGCCCTGTGCTTCTCGGGAACCATGAGTGCAGAAAGGCAACTGGAGGGAAAT
CTGAGGCAGGACACGGTCAGACCTAGCCATGCTCAGGTTTGGCCTGATAAGTTTATCAGCCTGTGCCGGCTGCCA
ACTGTCAGCATGGAGCCTGGCAGGGGGCTGGGTGGGAGGACACTCCTGGGTCGATAGCGCCTTCCAAGTGTTATC
AGGGGCACCGTGGCCCAGAGCGCGGGAAACGGGTGAAACAGGAGAGAAAGAGACTTCCAACCACTTTACCCAAAG
AAAAGCACTGGGAGGGGAACCGAGGCCTCAGTGTTCCTGGACCCTGACCTGCTGTGAAACCAGCAGTCACAGCTG
AGTCCCAGGGAGGCACAGGCTGAGGACCCTGCCACAGACATCCAGAGGGAA

Gfi1b+13 mutants for Gfi (yellow) and Ebox (blue) were generated using standard recombinant DNA techniques using the primers listed below.

Gfi1b13_Gfimut_Fw	ggaattctgaggcaggacacggtcagacc
Gfi1b13_Gfimut_Rv	agaattccctccagttgcctttctgc

Gfi1b13_Eboxmut_Fw ttcctgcagttatcaggggcaccgtggc

Gfi1b13_Eboxmut_Rv taactgcaggaaggcgctatcgacccagg

The Gata (green) and Ets (purple/red) mutants were generated by GeneArt® Gene Synthesis (Gata and Ets1-2) or GeneArt® Strings™ (Ets3-5) from Life Technologies. The whole Gfi1b+13 enhancer fragment with the relevant point mutations was ordered and subsequently cloned into pGL2 promoter vector from Promega.

Gfi1b+13_Gata (generated by GeneArt® Gene Synthesis from Life Technologies):
1st GATA-GGTA, 2nd TATC-GATC, 3rd GATA-GGTA, 4th TATC-GATC

Gfi1b+13_Ets1-2 (generated by GeneArt® Gene Synthesis from Life Technologies):
1st: GAGGGAA – AAGCTTA 2nd: GCCTTCC – GCTAGCC

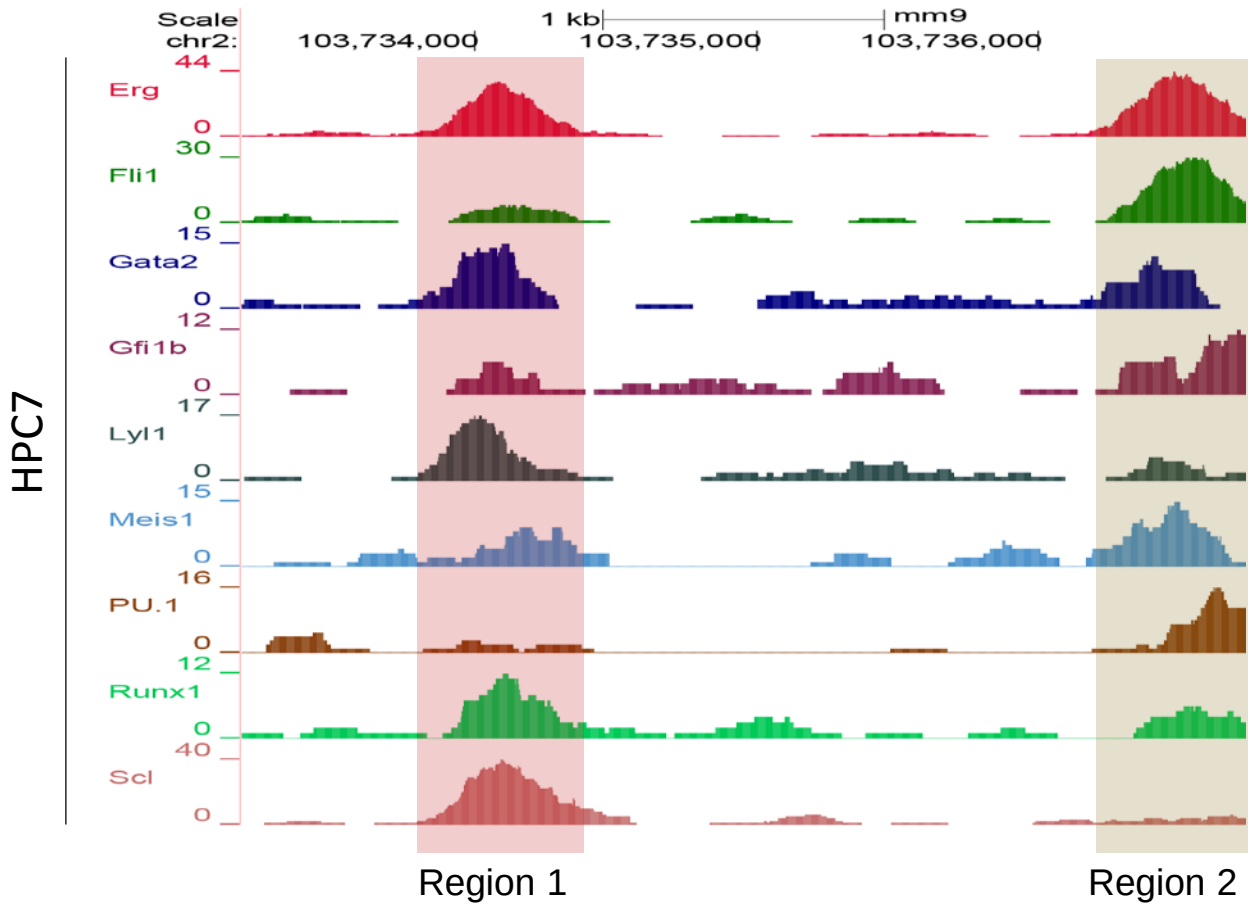
Gfi1b+13_Ets3-5 (generated by GeneArt® Strings™ from Life Technologies):
3rd: TTCC – TTGC 4th: GGGAAC – GACGTC 5th: TTCC - TACC

Supplementary Figure 4 *ConBind identifies conserved TFBSs within the Lmo2-75 enhancer.* ***A.*** *The previously described, hematopoietic active regulatory region for the Lmo2 gene (called Lmo2-75, (1)) is 3.5 kb long and is comprised of two sub-regions (region 1 marked in red, region two marked in yellow) that are bound by several TFs in the hematopoietic progenitor cell line HPC7.* ***B.*** *Luciferase reporter assays in stably transfected 416b cells, a myeloid progenitor cell line, reveal that both sub-regions of this enhancer are transcriptionally active on their own. Shown is the relative luciferase activity of the wild-type (wt) enhancer compared to an empty control vector.* ***C.*** *ConBind's algorithm results in aligning two ETS motifs, one Ebox and one Gata motif within sub-region 1, matching the TFs binding to this part of the enhancer. ConBind is able to find additional ETS and GATA motifs (six and four in total, respectively) within sub-region 2 of the Lmo2-75 enhancer.* ***D.*** *Manual identification of TFBSs within the two TF-bound sub-regions of the Lmo2-75 enhancer was performed as described in Figure 4A. Despite binding of Erg, Gata2, Lyl1, Runx1 and Scl (see A), no conserved TFBS could be observed within sub-region 1. In contrast four conserved ETS motifs, two conserved GATA sites and three conserved GFI1 motifs are located within sub-region 2 where binding of the ETS factors Erg, Fli1 and PU.1 as well as Gata2 can be seen (see A).*

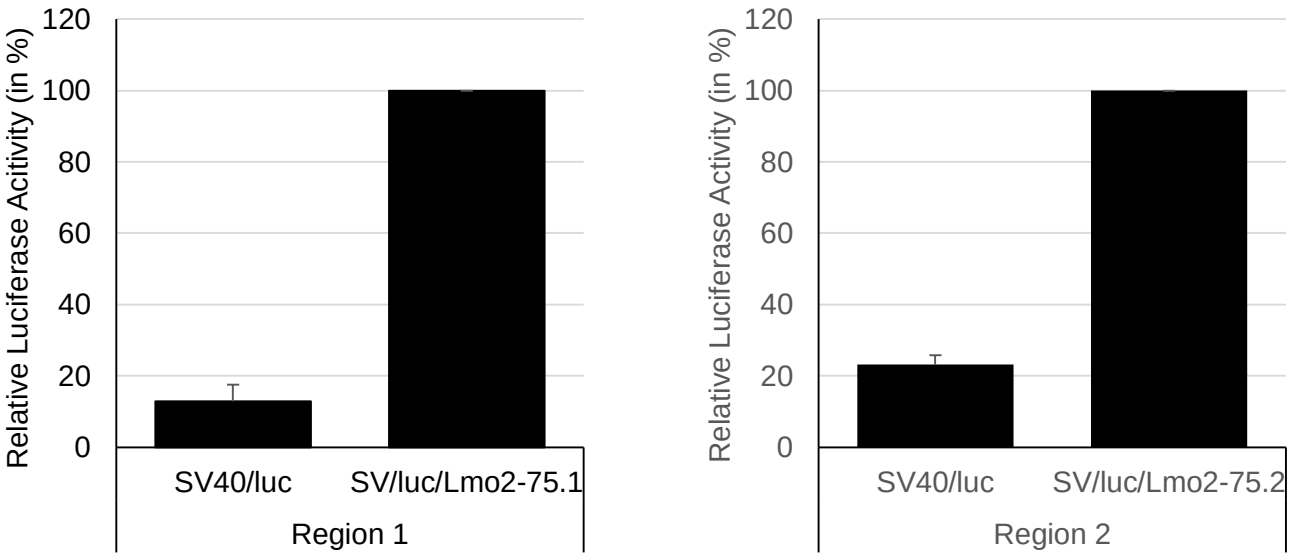
1. Landry, J.R., Bonadies, N., Kinston, S., Knezevic, K., Wilson, N.K., Oram, S.H., Janes, M., Piltz, S., Hammett, M., Carter, J. *et al.* (2009) Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood*, **113**, 5783-5792.

A

Lmo2-75: chr2:103733174-103736735



B



C

http://www.conbind.org/view/f6bc8728-e6ae-405b-b430-8e292016a3dc - 1

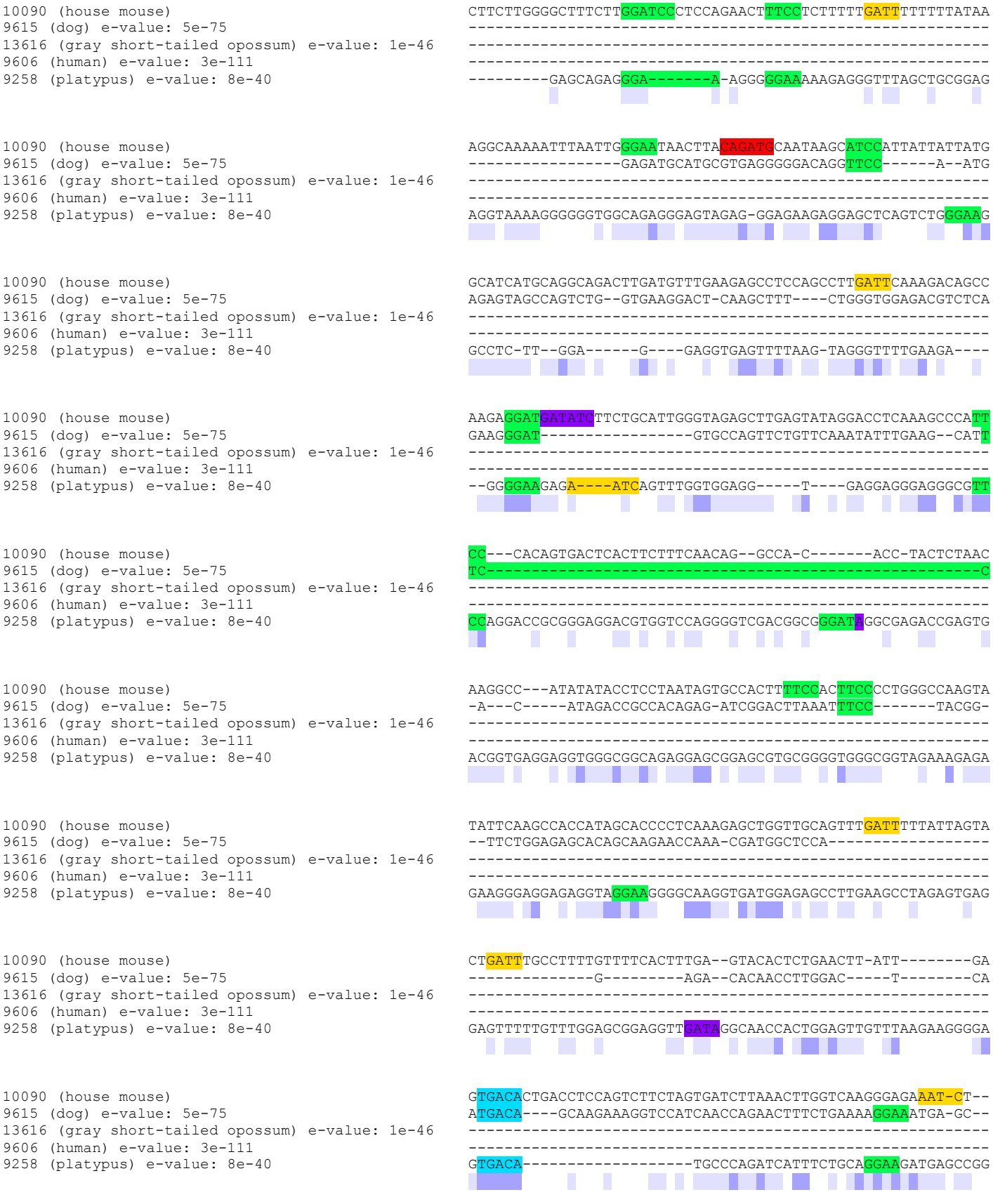
Legend

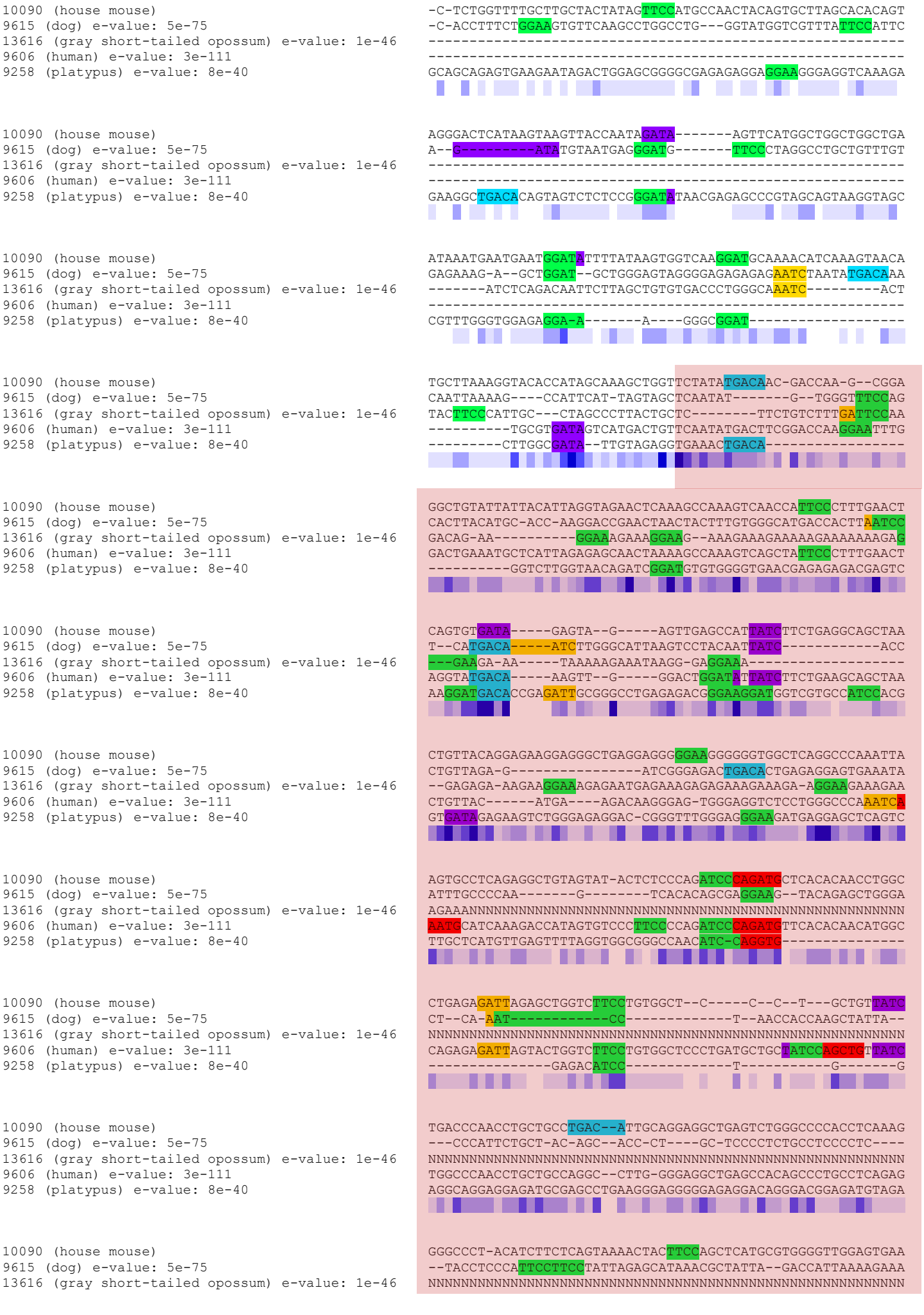
EBOX ETS GATA GFI1 MEIS RUNT

Alignment Quality: Low High

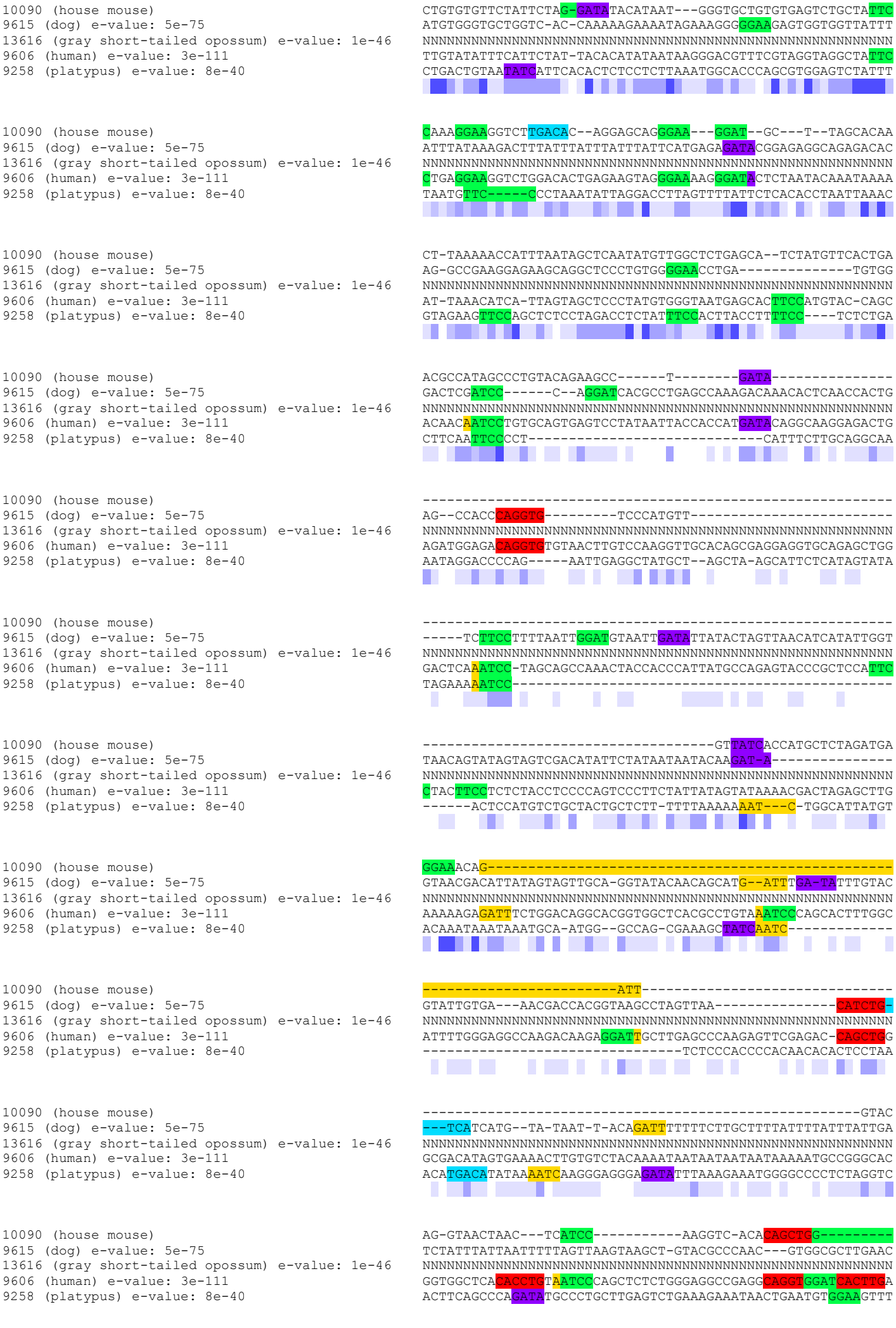
Lmo2-75

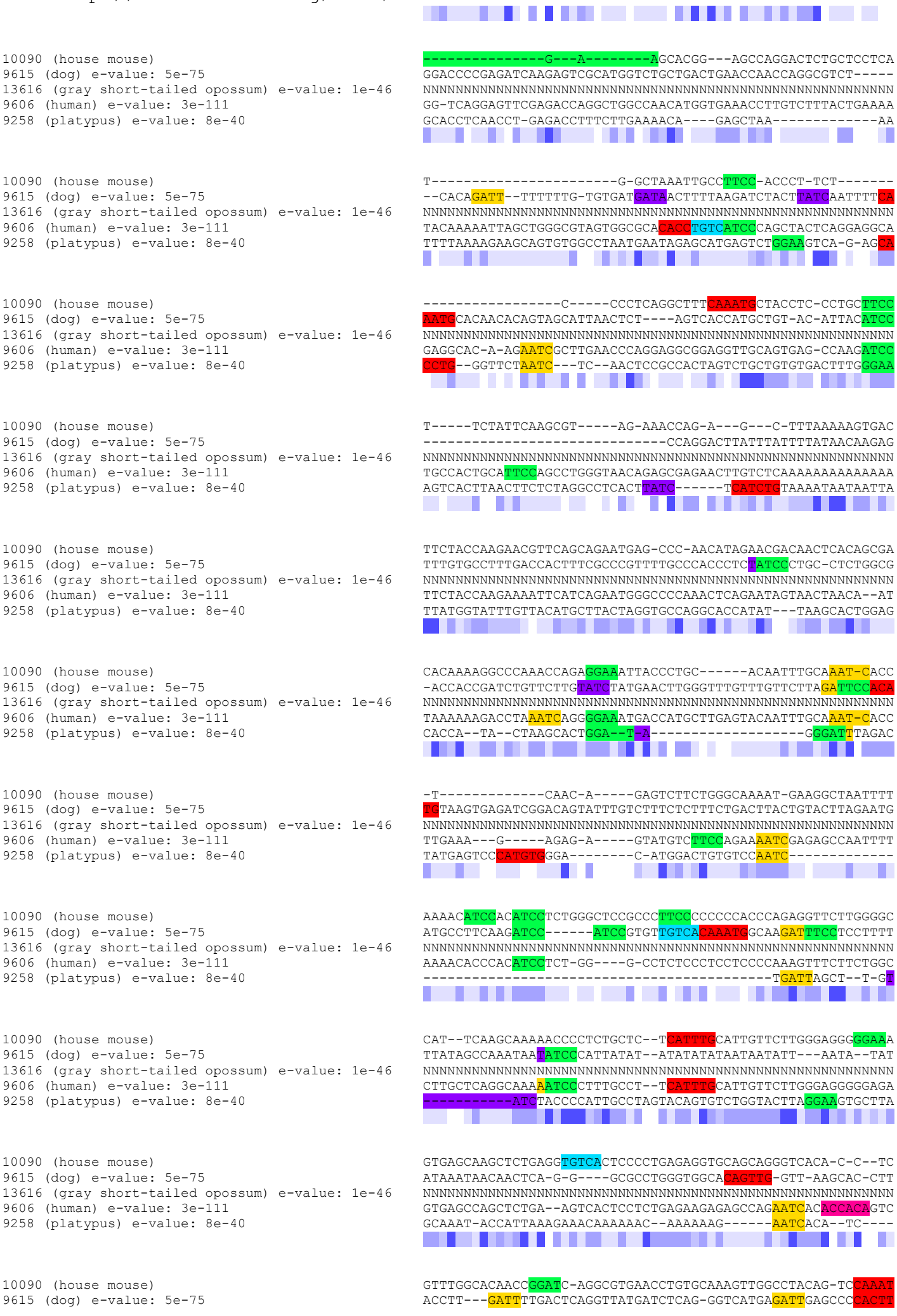
chr2:103733174-103736735

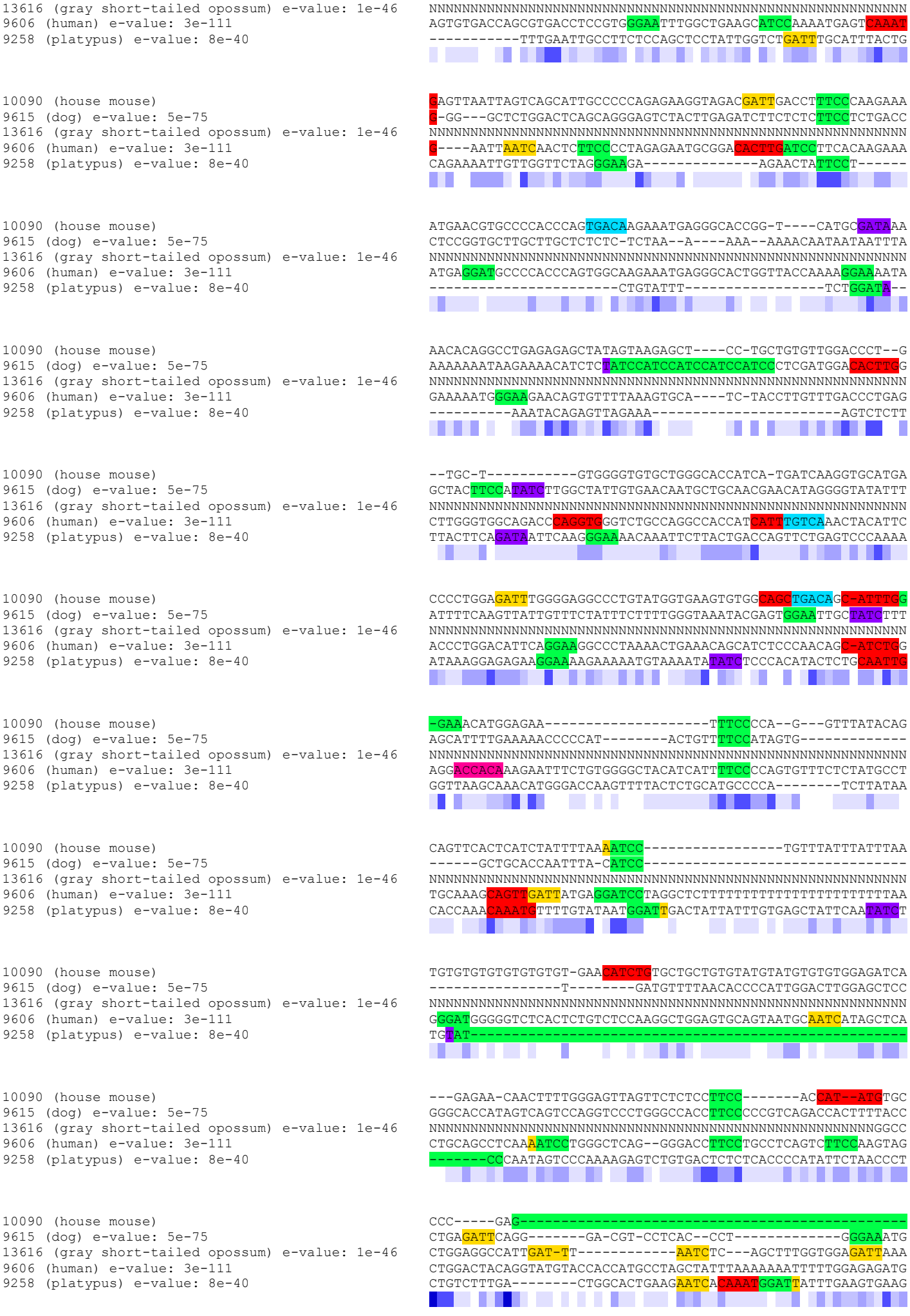


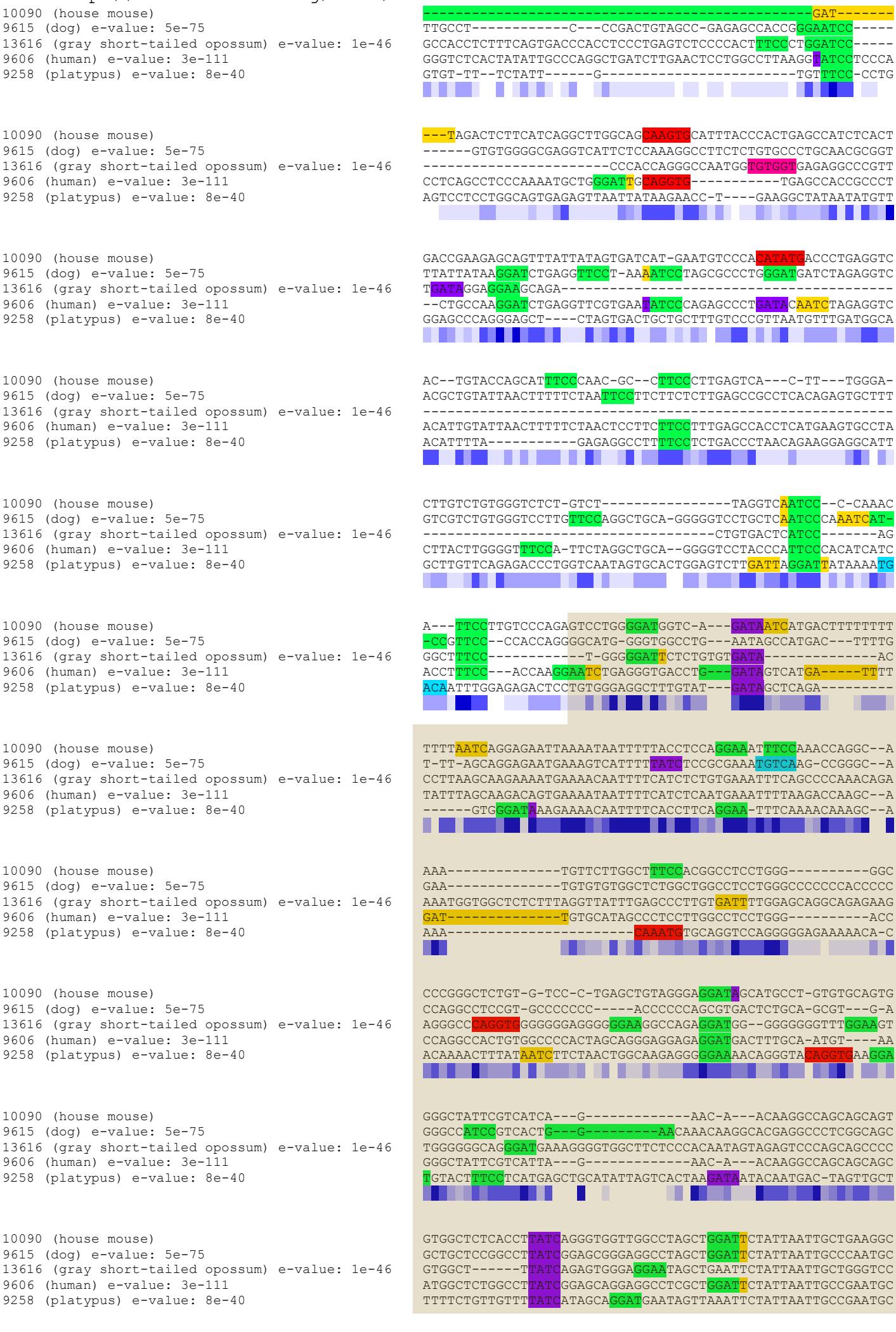

10090 (house mouse)
9615 (dog) e-value: 5e-75
13616 (gray short-tailed opossum) e-value: 1e-46
9606 (human) e-value: 3e-111
9258 (platypus) e-value: 8e-40

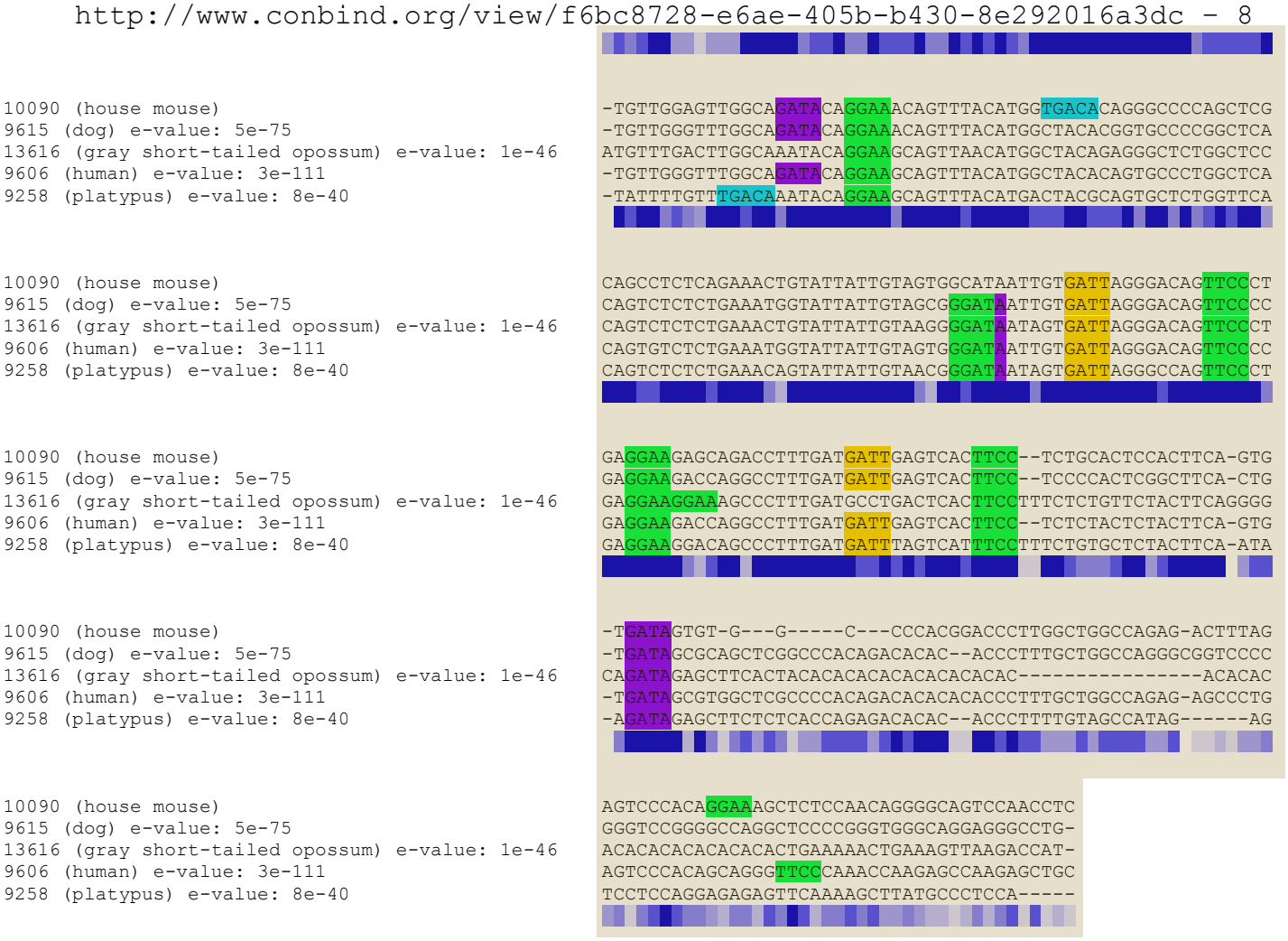
```
9606 (human) e-value: 3e-111                           GGGCCCTGTCATCTTTCCATTAAAGCTACCTCCA----TTACTTCTAGTGACGGAAAT
9258 (platypus) e-value: 8e-40                        TCTGCGTGTCATCTGCATAGAGATGGTAGTCAAAG------C----CATGAGAGCGAA

10090 (house mouse)                                    AGTGTCCCACTGACGATAGAGGAGGAGAGGATTGGTCTCTAGAATGACCTCCTGTCTG
9615 (dog) e-value: 5e-75                             TTTCTCCCA------AGAAAGTTCGTCAGAATGGGCCCCAACACAGAACAGCAACTAA
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           GTTCTACCAGCTACAGGGCATTGGGGTGTGCTTAGTCTCCAGAACCACCTCTCATTAA
9258 (platypus) e-value: 8e-40                        TGAGT-------TC-------ACCGAGGGAGTGAGTGTAAATGGAGAACAGAAGAGGG

10090 (house mouse)                                    TCACCTAACCCTCATGGTTTAGGTGCCACA--GG-AGTGGA--------G-------A
9615 (dog) e-value: 5e-75                             CAACAGAAAACGCCCAAATTGGGGGAAATTACTACATTTGAGTACAATTTGCAAATCA
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           TCAGCCACCTCCAACCATTTTGGTGCCATAGGGGCAGAGATTCACCTCCGTAACAAAA
9258 (platypus) e-value: 8e-40                        CCA-AGAACTGACCCTTGCTCGTTGCCCGTACCCCCTCTAA-TAA-TAT-T---AATA

10090 (house mouse)                                    -CTTAGAAATGTGCAACGTATTCCCTGCTAACTGGGTGAGCCTTCTGATGTCTTAGAT
9615 (dog) e-value: 5e-75                             TCT-TGGGA----GAGAGTCTTCCC---------------------------------
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           TATATGAAACTTGGATGGTGTTTCCTGTTAAACGAATGAGTCTTCTGATTTCTTTGAT
9258 (platypus) e-value: 8e-40                        ATGTTGGTATTTGTTAAGCATTTACTATGTGCAGAG-CACTGTTCTAAGCCCTGGGGT

10090 (house mouse)                                    CTCAGCGTCCCGTGAAATTGAGTGGGGATA---------A-CTTCCACAGTCTACGGC
9615 (dog) e-value: 5e-75                             ------------GCAAAGTGAGAGCCAATTTTAAAACACCCACATCCTC-----TG--
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           CTCACCCTCCTACCAGACAGA-AATTGATAAATGGGGACATGTTCCAATGAGAGTGAC
9258 (platypus) e-value: 8e-40                        -------------------------AGATA----------CAGGGTAATCAAGTTGTC

10090 (house mouse)                                    ACACCTCA-GGAAAGCGCTCTCAA-CTAT-TCCAAGTGAAC-AGCT---AAAGGGTGT
9615 (dog) e-value: 5e-75                             --GCCTCACTCCCTCCTC-CCCAAAG---GTTCATCTGGCCTAGCTCGGCAAAAA--T
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           CAGCCTCATGGACTGGACTCACAAGCATTCTGGATGGAGAC-ATCTCGGAAGGAATAT
9258 (platypus) e-value: 8e-40                        CCACATCAGACTCA-CAGTTAGTCCCATTTTACAGATGAGGTAACT------------

10090 (house mouse)                                    GCCAATTCTTTTAAGGGATTTGAAGGATTTCCATGTAGAACAAAAGCCAGA-C-GAGT
9615 (dog) e-value: 5e-75                             -----CCCTTTGCCTTCATTTGCATTG-TT---CTTGGGAGGAG-GAAAG---TGAGC
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           ACCAACTCTTTTGTAATATTTGATGAGTTTCCACACAGAAGCAGTGTTAGACCTAAAT
9258 (platypus) e-value: 8e-40                        -------------GAGGCACAGAGAAGTTGTGACTTGCCCCCAGTCACACAGCTGACA

10090 (house mouse)                                    TGTCTGTGATTCTGGGGCTCTCAGCAAGCAGTAAGTGATGATTACAGAGTGTTTGCCT
9615 (dog) e-value: 5e-75                             GAGCCCTG---CAGTCACTCCCCCTGAGAAGACAGTCA-GAATCAC-A-------CCT
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           TTCCTATGGTTCTGGAGAGCACAGCAAAAACTAAATGATGGCTCCAGAGCAACAGCAT
9258 (platypus) e-value: 8e-40                        AGTGGCAGA-GCCGGGGTTCAAACCCA--TGAC--CTCTGACTCCCAAGCCCGTGC-T

10090 (house mouse)                                    TTGACAAAAGTGAAAAGGAGAAATGTCTGCTAACCATAGCTGTCTGAACACTCCCTCC
9615 (dog) e-value: 5e-75                             TCAACTCAA-TCG---GCCGTGACTTCTTGGGA--ATTTGG--CTGAAGGGTCCCAAA
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           TTAATTCAAAT-AAAAGAAGAAATCTCTATCAACC--A-------GAATGATCTG---
9258 (platypus) e-value: 8e-40                        CTT--TCAACTG------------------AGCCACGCTGCTTCTCTAGGCTGTAAGC

10090 (house mouse)                                    CGACCCCCACTCCGAACCTTTCCTCCCCCCCCCCCCCGCCCCCCCCGCCAGCTGAATAC
9615 (dog) e-value: 5e-75                             CGAGTCAAACT---TAATTAGTCAACTTTTACCCCAGAGCATGTAGACACTTGA-TGC
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           --------------AA--------------------------AACGAAATGGACTAC
9258 (platypus) e-value: 8e-40                        TCGTTGTGGGCAAGGAATGTGTCTA-TTTTT--CTATTTTACTCTCCCAAGTGA----

10090 (house mouse)                                    CCAGCAGGGGATTGAGTTGCCTCACCCCATCTTTGGTGGAGTATGAGCTTGGCCTGAG
9615 (dog) e-value: 5e-75                             TTTGTAAGAAAATGAG-T---GCACCATACCCGT---------TGACAAG-----AAA
13616 (gray short-tailed opossum) e-value: 1e-46      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
9606 (human) e-value: 3e-111                           CTAGTTGAGTAGTGA---G--TCTCCT-GTCC------AAGCCTTACCTGGGTATGAG
9258 (platypus) e-value: 8e-40                        -TT--A--GTACAGTGCT-CTGCACACAGTAAGCTCTCAATAAATACGATTGACTGA-
```


10090 (house mouse)
9615 (dog) e-value: 5e-75
13616 (gray short-tailed opossum) e-value: 1e-46
9606 (human) e-value: 3e-111
9258 (platypus) e-value: 8e-40


10090 (house mouse)
9615 (dog) e-value: 5e-75
13616 (gray short-tailed opossum) e-value: 1e-46
9606 (human) e-value: 3e-111
9258 (platypus) e-value: 8e-40

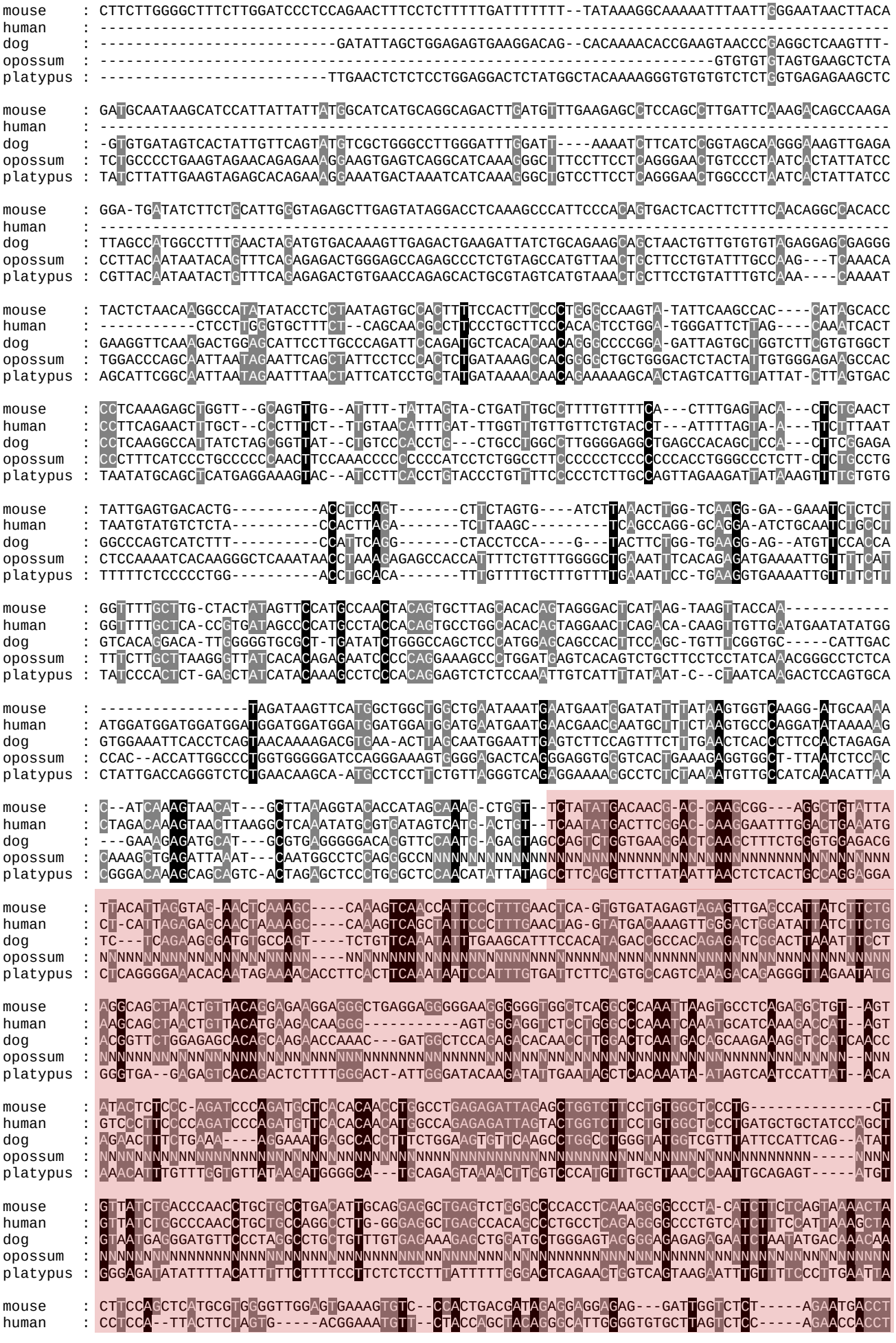

13616 (gray short-tailed opossum) e-value: 1e-46
9606 (human) e-value: 3e-111
9258 (platypus) e-value: 8e-40
10090 (house mouse)
9615 (dog) e-value: 5e-75

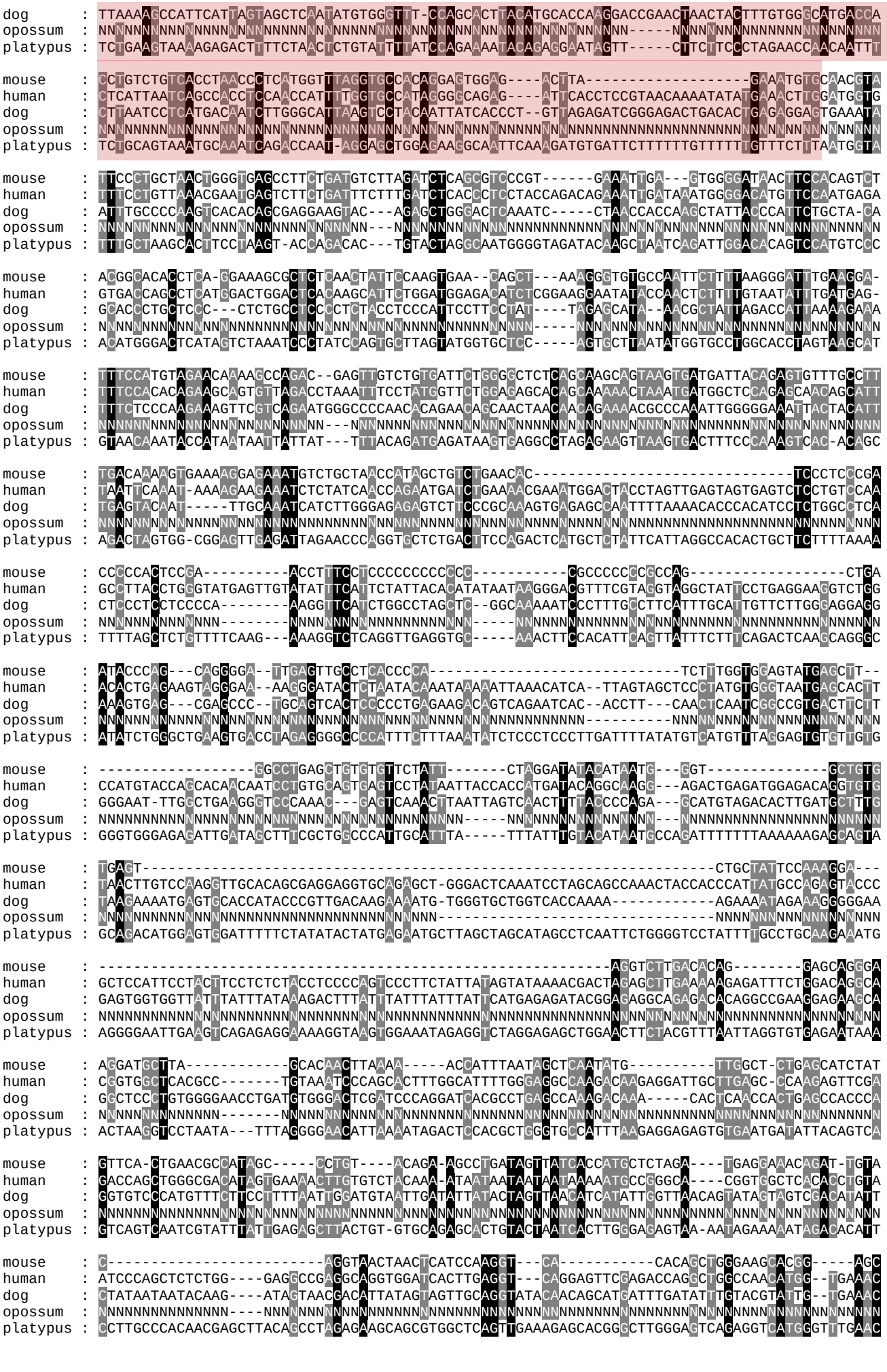

10090 (house mouse)
9615 (dog) e-value: 5e-75
13616 (gray short-tailed opossum) e-value: 1e-46
9606 (human) e-value: 3e-111
9258 (platypus) e-value: 8e-40

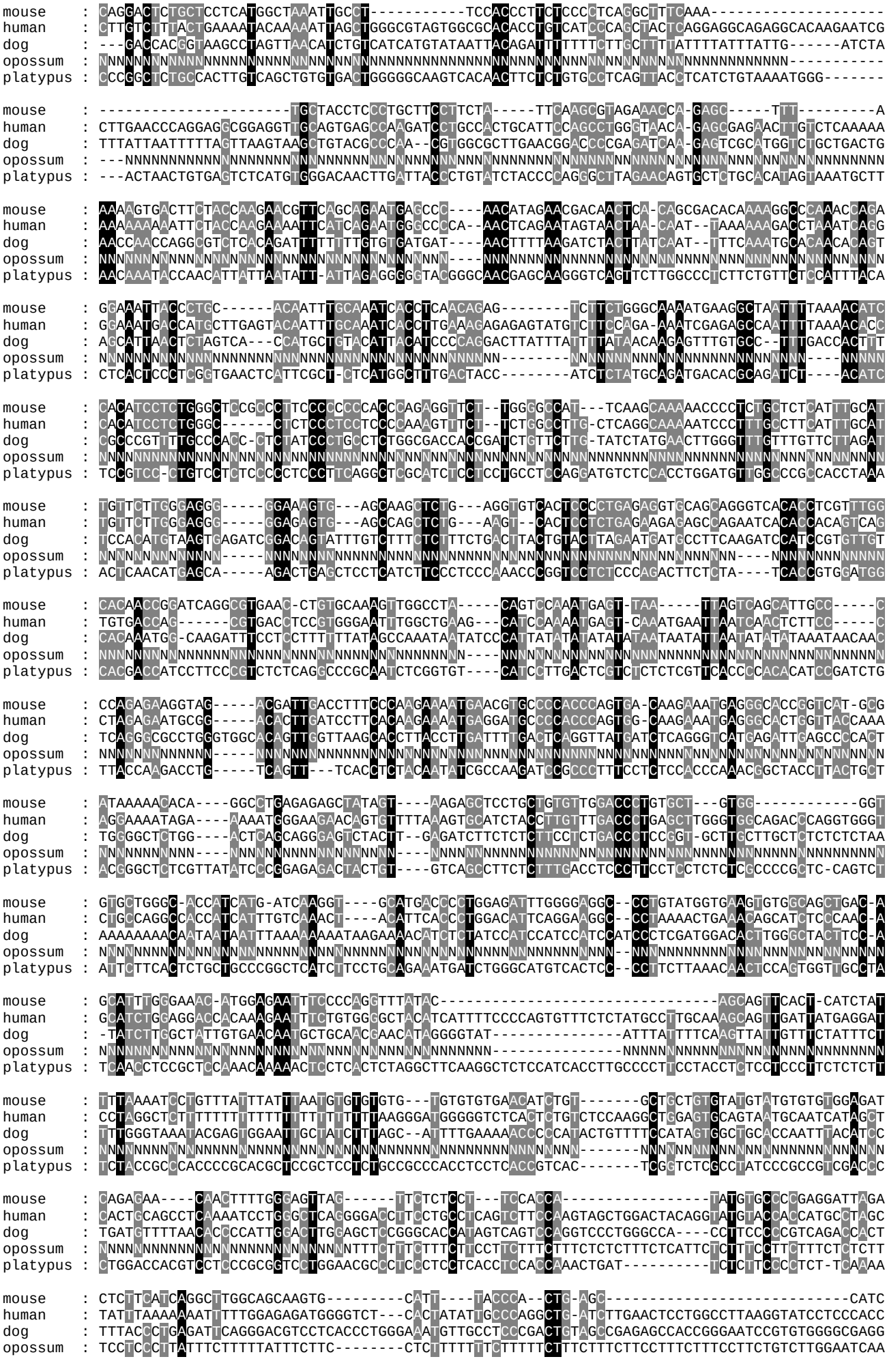

10090 (house mouse)
9615 (dog) e-value: 5e-75
13616 (gray short-tailed opossum) e-value: 1e-46
9606 (human) e-value: 3e-111
9258 (platypus) e-value: 8e-40
-TGTTGGAGTTGGCAGATACAGGAAACAGTTTACATGGTGACACAGGGCCCCAGCTCG
-TGTTGGGTTTGGCAGATACAGGAAACAGTTTACATGGCTACACGGTGCCCCGGCTCA
ATGTTTGACTTGGCAAATACAGGAAGCAGTTAACATGGCTACAGAGGGCTCTGGCTCC
-TGTTGGGTTTGGCAGATACAGGAAGCAGTTTACATGGCTACACAGTGCCCTGGCTCA
-TATTTTGTTTGACAAATACAGGAAGCAGTTTACATGACTACGCAGTGCTCTGGTTCA
CAGCCTCTCAGAAACTGTATTATTGTAGTGGCATAATTGTGATTAGGGACAGTTCCCT
CAGTCTCTCTGAAATGGTATTATTGTAGCGGGATAATTGTGATTAGGGACAGTTCCCC
CAGTCTCTCTGAAACTGTATTATTGTAAGGGGATAATAGTGATTAGGGACAGTTCCCT
CAGTGTCTCTGAAATGGTATTATTGTAGTGGGATAATTGTGATTAGGGACAGTTCCCC
CAGTCTCTCTGAAACAGTATTATTGTAACGGGATAATAGTGATTAGGGCCAGTTCCCT
GAGGAAGAGCAGACCTTTGATGATTGAGTCACTTCC--TCTGCACTCCACTTCA-GTG
GAGGAAGACCAGGCCTTTGATGATTGAGTCACTTCC--TCCCCACTCGGCTTCA-CTG
GAGGAAGGAAAGCCCTTTGATGCCTGACTCACTTCCTTTCTCTGTTCTACTTCAGGGG
GAGGAAGACCAGGCCTTTGATGATTGAGTCACTTCC--TCTCTACTCTACTTCA-GTG
GAGGAAGGACAGCCCTTTGATGATTTAGTCATTTCCTTTCTGTGCTCTACTTCA-ATA
-TGATAGTGT-G---G-----C---CCCACGGACCCTTGGCTGGCCAGAG-ACTTTAG
-TGATAGCGCAGCTCGGCCCACAGACACAC--ACCCTTTGCTGGCCAGGGCGGTCCCC
CAGATAGAGCTTCACTACACACACACACACACACAC----------------ACACAC
-TGATAGCGTGGCTCGCCCCACAGACACACACACCCTTTGCTGGCCAGAG-AGCCCTG
-AGATAGAGCTTCTCTCACCAGAGACACAC--ACCCTTTTGTAGCCATAG------AG
AGTCCCACAGGAAAGCTCTCCAACAGGGGCAGTCCAACCTC
GGGTCCGGGGCCAGGCTCCCCGGGTGGGCAGGAGGGCCTG-
ACACACACACACACACTGAAAAACTGAAAGTTAAGACCAT-
AGTCCCACAGCAGGGTTCCCAAACCAAGAGCCAAGAGCTGC
TCCTCCAGGAGAGAGTTCAAAAGCTTATGCCCTCCA-----

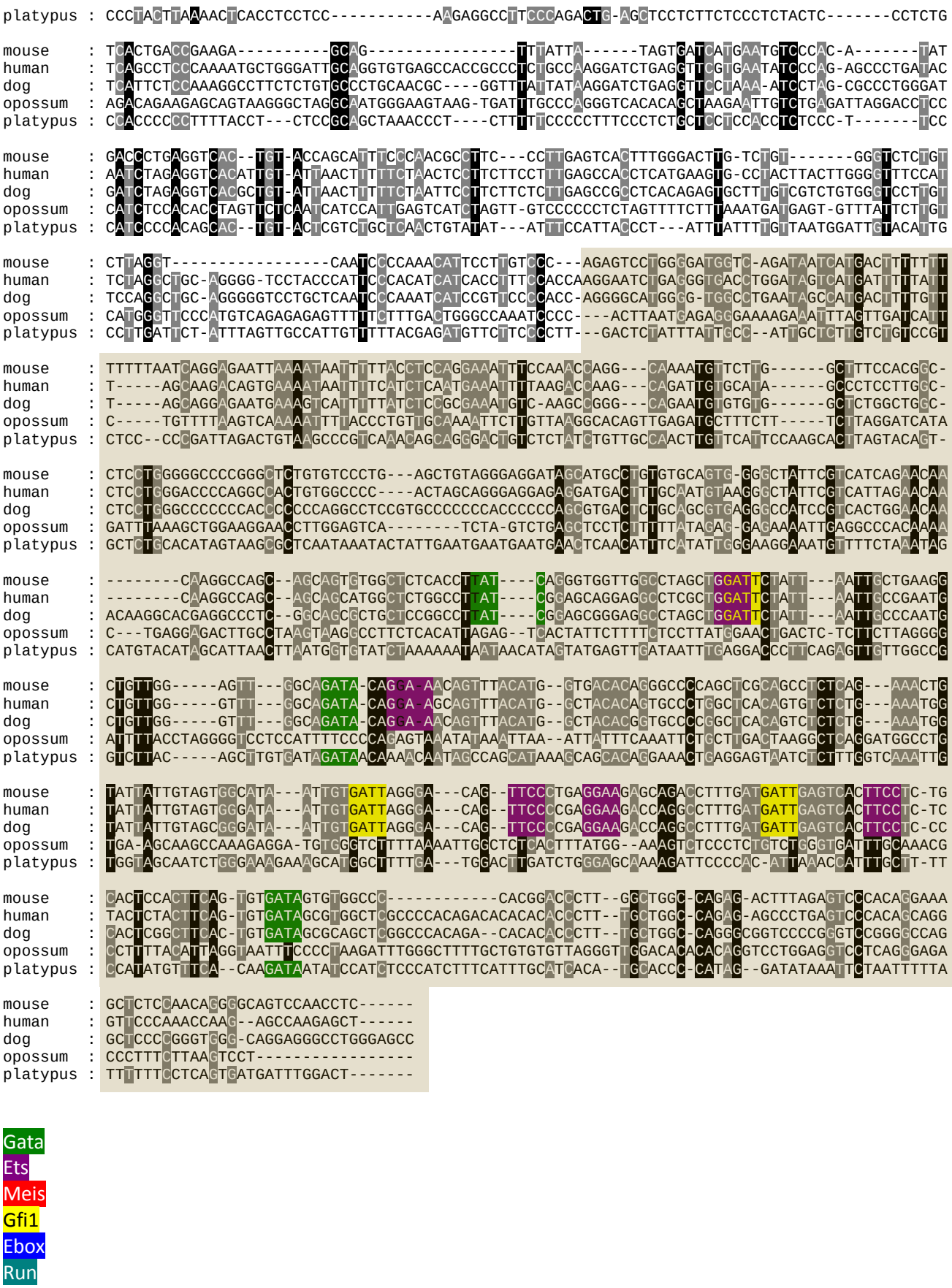
D

Lmo2 -75kb: Manual alignment




dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus
mouse
human
dog
opossum
platypus




platypus : CCCTACTTAAAACTCACCTCCTCC----------AAGAGGCCTTCCCAGACTG-AGCTCCTCTTCTCCCTCTACTC-------CCTCTG
mouse : TGACTGACCGAAGA----------GCAG----------TTTATTA------TAGTGATCATGAATGTCCCAC-A-------TAT
human : TCAGCCTCCCAAAATGCTGGGATTGCAGGTGTGAGCCACCGCCCTCTGCCAAGGATCTGAGGTTCGTGAATATCCCAG-AGCCCTGATAC
dog : TCATTCTCCAAAGGCCTTCTCTGTGCCCTGCAACGC----GGTTTATTATAAGGATCTGAGGTTCCTAAA-ATCCTAG-CGCCCTGGGAT
opossum : AGACAGAAGAGCAGTAAGGGCTAGGCAATGGGAAGTAAG-TGATTTGCCCAGGGTCACACAGCTAAGAATTGTCTGAGATTAGGACCTCC
platypus : CCACCCCCCTTTTACCT---CTCCGCAGCTAAACCCT----CTTTTTCCCCCTTTCCCTCTGCTCCTCCACCTCTCCC-T-------TCC
mouse : GACCCTGAGGTCAC--TGT-ACCAGCATTTCCCAACGCCTTC---CCTTGAGTCACTTTGGGACTTG-TCTGT-------GGGTCTCTGT
human : AATCTAGAGGTCACATTGT-ATTAACTTTTTCTAACTCCTTCTTCCTTTGAGCCACCTCATGAAGTG-CCTACTTACTTGGGGTTTCCAT
dog : GATCTAGAGGTCACGCTGT-ATTAACTTTTTCTAATTCCTTCTTCTCTTGAGCCGCCTCACAGAGTGCTTTGTCGTCTGTGGGTCCTTGT
opossum : CATCTCCACACCTAGTTCTCAATCATCCATTGAGTCATCTAGTT-GTCCCCCCTCTAGTTTTCTTTAAATGATGAGT-GTTTATTCTTGT
platypus : CATCCCCACAGCAC--TGT-ACTCGTCTGCTCAACTGTATAT---ATTTCCATTACCCT---ATTTATTTTGTTAATGGATTGTACATTG
mouse : CTTAGGT----------------CAATCCCCAAACATTCCTTGTCCC---AGAGTCCTGGGGATGGTC-AGATAATCATGACTTTTTTT
human : TCTAGGCTGC-AGGGG-TCCTACCCATTCCCACATCATCACCTTTCCACCAAGGAATCTGACGGTGACCTGGATAGTCATGATTTTTATT
dog : TCCAGGCTGC-AGGGGGTCCTGCTCAATCCCAAATCATCCGTTCCCCACC-AGGGGCATGGGG-TGGCCTGAATAGCCATGACTTTTGTT
opossum : CATGGGTTCCCATGTCAGAGAGAGTTTTTCTTTGACTGGGCCAAATCCCC----ACTTAATGAGAGGGAAAAGAAATTTAGTTGATCATT
platypus : CCTTGATTCT-ATTTAGTTGCCATTGTTTTTACGAGATGTTCTTCCCCTT---GACTCTATTTATTGCC--ATTGCTCTTGTCTGTCCGT
mouse : TTTTTAATCAGGAGAATTAAAATAATTTTTACCTCCAGGAAATTTCCAAACCAGG---CAAAATGTTCTTG------GCTTTCCACGGC-
human : T-----AGCAAGACAGTGAAAATAATTTTCATCTCAATGAAATTTTAAGACCAAG---CAGATTGTGCATA------GCCCTCCTTGGC-
dog : T-----AGCAGGAGAATGAAAGTCATTTTTATCTCCGCGAAATGTC-AAGCCGGG---CAGAATGTGTGTG------GCTCTGGCTGGC-
opossum : C-----TGTTTTAAGTCAAAAATTTTACCCTGTTGCAAAATTCTTGTTAAGGCACAGTTGAGATGCTTTCTT-----TCTTAGGATCATA
platypus : CTCC--CCCGATTAGACTGTAAGCCCGTCAAACAGCAGGGACTGTCTCTATCTGTTGCCAACTTGTTCATTCCAAGCACTTAGTACAGT-
mouse : CTCCTGGGGGCCCCGGGCTCTGTGTCCCTG---AGCTGTAGGGAGGATAGCATGCCTGTGTGCAGTG-GGGCTATTCGTCATCAGAACAA
human : CTCCTGGGACCCCAGGCCACTGTGGCCCC----ACTAGCAGGGAGGAGAGGATGACTTTGCAATGTAAGGGCTATTCGTCATTAGAACAA
dog : CTCCTGGGCCCCCCCACCCCCCAGGCCTCCGTGCCCCCCCACCCCCCAGCGTGACTCTGCAGCGTGAGGGCCATCCGTCACTGGAACAA
opossum : GATTTAAAGCTGGAAGGAACCTTGGAGTCA--------TCTA-GTCTGAGCTCCTCTTTTTATAGAG-GAGAAAATTGAGGCCCACAAAA
platypus : GCTCTGCACATAGTAAGCGCTCAATAAATACTATTGAATGAATGAATGAACTCAACATTTCATATTGGGAAGGAAATGTTTTCTAAATAG
mouse : --------CAAGGCCAGC--AGCAGTGTGGCTCTCACCTTTAT----CAGGGTGGTTGGCCTAGCTGGATTCTATT---AATTCCTGAAGG
human : --------CAAGGCCAGC--AGCAGCATGGCTCTGGCCTTTAT----CGCAGCAGGAGGCCTCGCTGGATTCTATT---AATTCCGAATG
dog : ACAAGGCACGAGGCCCTC--GGCAGCGCTGCTCCGGCCTTTAT----CGCAGCGGGAGGCCTAGCTGGATTCTATT---AATTCCCCAATG
opossum : C---TGAGGAGACTTGCCTAAGTAAGCCTTCTCACATTAGAG--TCACTATTCTTTTCTCCTTATGGAACTGACTC-TCTTCTTAGGGG
platypus : CATGTACATAGCATTAACTTAATGGTGTATCTAAAAAATAATAACATAGTATGAGTTGATAATTTGAGGACCCTTCAGAGTTGTTGGCCG
mouse : CTGTTGG-----AGTT---GGCAGATA-CAGGA-AACAGTTTACATG--GTGACACAGGGCCCCAGCTCGCAGCCTCTCAG---AAACTG
human : CTGTTGG-----GTTT---GGCAGATA-CAGGA-AGCAGTTTACATG--GCTACACAGTGCCCCTGGCTCACAGTGTCTCTG---AAATGG
dog : CTGTTGG-----GTTT---GGCAGATA-CAGGA-AACAGTTTACATG--GCTACACGGTGCCCCGGCTCACAGTCTCTCTG---AAATGG
opossum : ATTTTACCTAGGGGTCCTCCATTTTCCCCAGAGTAAATATAAATTAA--ATTATTTCAAATTCTGCTTGAGTAAGGCTCAGGATGGCCTG
platypus : GTCTTAC-----AGCTTGTGATAGATAACAAAACAATAGCCAGCATAAAGCAGCACAGGAAACTGAGGAGTAATCTCTTTGGTCAAATTG
mouse : TATTATTGTAGTGGCATA---ATTGTGATTAGGGA---CAG--TTCCCTGAGGAAGAGCAGACCTTTGATGATTGAGTCACTTCCTTC-TG
human : TATTATTGTAGTGGGATA---ATTGTGATTAGGGA---CAG--TTCCCCGAGGAAGACCAGGCCTTTGATGATTGAGTCACTTCCTTC-TC
dog : TATTATTGTAGCGGGATA---ATTGTGATTAGGGA---CAG--TTCCCCGAGGAAGACCAGGCCTTTGATGATTGAGTCACTTCCTC-CC
opossum : TGA-AGCAAGCCAAAGAGGA-TGTCGGTCTTTTAAAATTGGCTCTCACTTTATGG--AAAGTCTCCCTCTGTCTGGGTGATTTGCAAACG
platypus : TGGTAGCAATCTGGGAAGAAGCATGGCTTTTGA---TGGACTTGATCTGGAGCAAAAGATTCCCCAC-ATTAAACCATTTGCTT-TT
mouse : CACTCCACTTCAG-TGTGATAGTGTGGCCC----------CACGGACCCTT--GGCTGGC-CAGAG-ACTTTAGAGTCCCACAGGAAA
human : TACTCTACTTCAG-TGTGATAGCGTGGCTCGCCCCACAGACACACACACCCTT--TGCTGGC-CAGAG-AGCCCTGAGTCCCACAGCAGG
dog : CACTCGGCTTCAC-TGTGATAGCGCAGCTCGGCCCACAGA--CACACACCCTT--TGCTGGC-CAGGGCGGTCCCCGGGTCCGGGGCCAG
opossum : CCTTTTACATTAGGTAATTTCCCCTAAGATTTGGGCTTTTGCTGTGTGTTAGGGTTGGACACACACAGGTCCTGGAGGTCCTCAGGGAGA
platypus : CCATATGTTTCA--CAAGATAATATCCATCTCCCATCTTTCATTTGCATCACA--TGCACCC-CATAG--GATATAAATTTCTAATTTTTA
mouse : GCTCTCCAACAGGGCAGTCCAACCTC------
human : GTTCCCAAACCAAG--AGCCAAGAGCT------
dog : GCTCCCCGGGTGGG-CAGGAGGGCCTGGGAGCC
opossum : CCCTTTCTTAAGTCCT----------
platypus : TTTTTTCCTCAGTCATGATTTGGACT-------
Gata
Ets
Meis
Gfi1
Ebox
Run