

SUPPLEMENTARY INFORMATION

Targeting mammalian intron-encoded box C/D 2'-O-methylation guide RNAs into the Cajal body

Aline Marnef, Patrica Richard, Natalia Pinzón and Tamás Kiss

Results

To identify evolutionarily conserved and therefore, functionally important sequences in vertebrate box C/D scaRNAs, the coding genes of the intron-encoded mgU2-47, U90, mgU2-19/30, and the independently transcribed mgU2-25/61 and mgU12-22/U4-8 scaRNAs were identified in mammalian, bird, lizard, turtle, frog and/or fish genomes by BLAT search (UCSC Genome Browser). The predicted RNA sequences were aligned to maximum homology by using the ClustalW2 algorithm developed for multiple sequence alignment (EMBL-EBI). In addition to the conserved C/D and C'/D' boxes and the RNA-specific target recognition sequences, we detected long GU- and AU-rich internal repeat sequences in the intronic mgU2-47, U90 and mgU2-19/30 scaRNAs (Figures S1, S2 and S3). The 5'-terminal regions of the independently transcribed mgU2-25/61 scaRNAs are also enriched in GU dinucleotide repeat sequences (Figure S4). In contrast, compilation and comparison of the Pol II-transcribed mgU12-22-U4-8 box C/D scaRNA sequences failed to identify GU-rich internal sequences (Figure S5).

	box C	tSL	
cat	AGCAAAGT GATGAG -CGTTGCTGGCTGGAG-CCCCAAAG----AGGCACACG- TGTGTGTGTGTGTGTATGTATGTATATATATATCTGTGCCCTTGTCTGTGCATGCACACGTATGT		113
armadillo	AGCAAAGT GATGAG -CTTTACTGGCTGGAG-CCCCAAAG----AGGCACACG- TGTGTGTGTGTGTGTGTCTGTGTTTGTATATATGTATGTCTTGTGTAT -----GCACAGGTGTTT		108
dog	AGCAAAGT GATGAG -CATTATTGGCTAGAG-CCCCAAAG----AGGCACATG- TGTGTATGTGTGTGTATGTAATGTGTCTGTATCCTTGTCTGTGCATG -----CACATGTATGT		105
human	AGCAAAGT GATGAG -TAATACTGGCTGGAG-CCCCAAAG----AGGCACGTG- TGTGTGTTTGTGTGTGTGTATATGCTTGTTCAGTGCATGCACGTGTATGT -----		97
rabbit	AGCAAAT GATGAG -CCACACTGGCTGGAT-CCCCAAAG----AGGCACGTG- TGTGTGTGTGTGTGTGTGTGTGTCTGTCCGTGCCCTCCTTGTCTGTGCGCTCCC-CGTGTGT		112
mouse	AGCA ATGATGAA -CAATACTG-CTGGAG-CCACAGAG----AGGCACAAG- TGTGTGTGCCAGATTGCCTGTGTGTATGTGTGTCTTGTCTGTGCATG -----CACACGTGAGT		104
opossum	AGCAAAT GATGGA -TATTATTGGTTGGAG-CCCTAAAG----AGGCA-ATA- TGAATGTATATGAATGACTGTATGTG -----CGCGCACT TGTGC		84
chicken	AGCAAAGT GATGAA -TAATATTTCTTGAG-CCAAAAG----AGAGAGAGT- TGTGTGTGTGTGGTTGTATGA -----		58
zebrafinch	AGCAAAT GATGAG -AAACACTTGCATGAG-CCGAAGTG----AGTGAGTTG- TGTGTGTGTGTGCAGTGGTGTGGTTGTTTGA -----		67
Xenopus	GGCAAAT GATGAG -ATCATCAGACTTGACTCCCCACAG---- TGTGACTCG-TG-GCACTTGTGTGTGTTTGTGTGCATGTTTGTGCATGT -----		71
lamprey	AGCAGG ATGATGA AAAACTTTTTCTTGAG-CCCCAAAATAAAATTCACACGTTGTGTATTTGTTT- TGTAATTGTGCGTA -----ACACGTATTC		89
		antisense box D	
cat	CTGGGAGTTT GAGTGTATGTGTGACTGGTGGTAGGGAACAAGTTAC ---GTGCCCTTC-TGTT---GGACCATGACAGT CAAACTGATAAGATCTGATTGCT		207
armadillo	CTGGGAGTACAAATG--TGTGTGACTGATTGTAGGAAATACTTAAT--GTGCCCTTC-TGTT---GGACTGTGACAGT CAAACTGATAAGATCTGATTGCT		201
dog	CTGGGAGTACGAATG--TGTCTGACTGATTATAGGAACTAGTCTAT--GTGCCCTTC-TGTT---GGACGATGACAGT CAAACTGATAAGATCTGATTGCT		198
human	CTGGGAGTACAAATG--GGTGCGACTGGTTGTAGGAACTAGCTAT--GTGCCCTTC-TATT---AGGCCATGACAGT CAAACTGATAAGATCTGATTGCT		189
rabbit	CTGG--CTATGCGTG---TGTGTCTGGTTATAGGAACTAGCTGT---GTGCCCTTC-TGTT---GGGCCATGACAGT CAAACTGATAAGATCTGATTGCT		200
mouse	CTGGGAGGATGGATG--TGTGTGACTGGTGGTAGGGACCTAGCCCT---GTGCCATC-TTCT---GGGCTGTGACAGT CAAACTGATAAGATCTGATTGCT		196
opossum	CTGGATGTATGACTG-GAGTATGAATGAGGATAAGGACCTAGTTAT---TTGCCCTTC-TTTTATAGGGCCATGATATTC CAAACTGATAAGATCTGATTGCT		180
chicken	-----CAG-----GGTGTTTGTGAG-----AATTGT---GTGCCCTTA-TAAT--TTGGTGGTGCAGCT CAAACTGATAAGATCTGATTGCT		140
zebrafinch	-----GAG-----AACTGT-----GTGCTTA-TAAT--TGGGCAGTGCATAGCT CAAACTGATAAGATCTGATTGCT		137
Xenopus	-----GGGTGACTGGTTGAGTG----TGAGTGAG--CACCTTC-ACATG--GGTTCCTGAGAGCT CAAACTGATAAGATCTGATTGCT		159
lamprey	ACGTTTGTTCATGTA--CACGTGTTGTTTATATGGTTCTGTTTTAATGTTGCACACGTGTGAACGGGCAGTGGGAAGGA CAAACTGATAAGATCTGACTGCT		187

Figure S1. Sequence alignment of vertebrate mgU2-47 RNAs. The conserved box C and D motifs and the antisense sequences are indicated in red and blue respectively. The RU-rich internal sequences are highlighted in yellow. The experimentally defined CB-localization stem-loop (tSL) of human mgU2-47 is shaded in blue.



Figure S2. Sequence alignment of vertebrate U90 RNAs. The conserved 5'-terminal C and 3'-terminal D boxes and the putative internal box C' and D' motifs are indicated in red. The predicted antisense sequences are in blue. The RU-rich internal sequences are highlighted in yellow. The experimentally defined CB-localization stem-loop (tSL) of human U90 is shaded in blue.

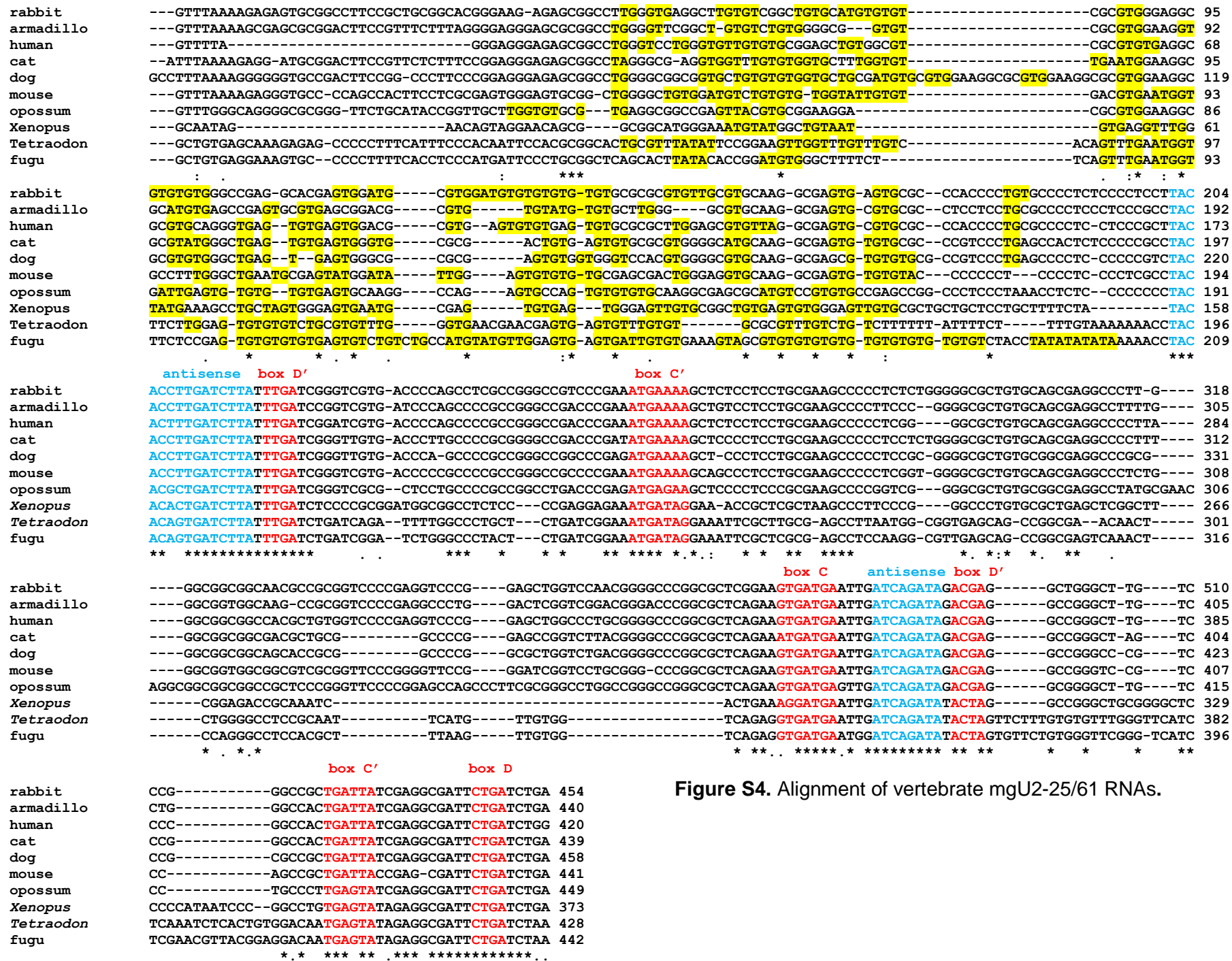


Figure S4. Alignment of vertebrate mgU2-25/61 RNAs.

```

human      ---AGAGGCTTGGGCCGCCG---AGCTGG-----ACCCGGACCCG-----TTTTGGG-TACTGTACTGGG--GCAGGGCA---GAGAGG-----TGGGCG----- 74
cat        ---CTGGGTTTGAGAGAGAC---CCTAGG-----ACTGGAAGCTA-----GTTGGCG-GGCCCGAGTGGGGTGGTGGGGT---GGGAGG-----TTGGCG----- 76
armadillo  ---AGATCAGTGGGCAGAG---CCTAGA-----TCTCTGCGCCG-----GACGAAGGAGCGTGGTGGCGGAGGCATGGT--GGGTGG-----GAGCGAAGAGG 85
rabbit     ---ATAGTATTTCTCAATGGTGACACTTGG--GAATTTGCAGAAGCTG-----CTTTGGAGTTCAGGTGGCG--GCCTGGGC-GCGGGGCGAG-----GCCGGCGGG 90
dog        ---AGAGGGCGGGCGCCGGGATTTGGCTGGGCGCCCGTGCGGGGCGGGGGCGGGCGGGCTGCGCGGAGGGGGCGAGGCTTAGCTGCGCGGGGAGAGGGGCGGGAGGA--- 114
mouse      ---GGTGGGATTGGCTGCCCTTGCCTGCGCAGCCTTCTGGCGGCGG-----CCTGTGCTAGTTCGG--AGCGG-GGCCGTGGC---GGGTAG-----CTGGGCG--- 88
opossum    ---CGGGCTGGGACAGTG--ATAACT--TCGGGAGCG-----GTTGCTTTGTCTCAGAT--CGCTCGGC-----GGTCT-----TTGGGGG--- 74
lizard     --ATGGTCCCCAAACGAAAGCATTCTTTGG-----GTGGTGGGCGG--TAGTGTTTGGGAAGGACATTGCGAGGGGT--TGGGCTG-----CATATT----- 81
turtle     ---TAAAGCTTTATATCTGTTAAACAC---AAATGTCAA-----CATCA--CATATCAGAATATGCAAAGTAA---ATAT----- 66
Coelacanth GTCTAAGGGTTTTTTTATTATTATTTTGG-----TAACTTTTCAG-----TTTTAGAAGTGTTTTACAATAACAATA---ATTTA----- 75

```

```

human      ---CGAGTGGGGTGCCTGTG---ATTGTAGTAGGCTAGGG-CGCTTTCGGGTC--CCCATTGCAGC-----CCCCGGATGAGCCCGCAGTATTTTCCTTATATGATCAGGTCCCA 175
cat        ---TCCGAGGGAGTGTAGTG---AGTGTCTTAGGGCAGGG-ATGCCCGGAGTC--CCGAAATAGC-----CC-CGATGAGCTCGCAGTATTTTCCTTATATGATCAGGCCCG 176
armadillo  TTGGCTTCGGGGGTGTAGCG---AACGTAGTAGGTTAGGGGCGCCGGGGATC--CCTGATCCAGC-----CCCTGGATGAGCCCTCAAATATTTTCCTTATATGATCAGGCCCTTT 190
rabbit     ---CGGTGGGGTGCCTGTG---AGGGTTG-TAGGGTAGGG-TGCGCCCGGATC--CCAAGTGCAGC-----CCCCGCATGAGCCCGCAGTGTTTTCCTTATATGATCAGGCCCG 190
dog        ---CCGGTGGGGGCGGTGGTC-GGGGTTGGAAGCGCGGGA-AGCCCCGAGTC--CCCAGTGTAGC-----CCCCTCGTGAGCTCGCGGTATTTTCCTTATATGATCAGGCTCCG 219
mouse      ---TCTAGAGAAGGGAAGGG---GAACGTGGTAGAGCAGGG-AGCGCTTGGGTC--CCCGAGGCAG-----CCCCGAGTGAAGCCGATGATTTTCCTTATATGATCAGGCCCT 189
opossum    ---AAGGAGGATTCTTTG---CAGTTTAGGAAGGGGAGAAGCGCCGAGTCGCCCAACGCAGC-----CCCCAAGTGA-CCACTATATTTTCCTTATATGATCAGGCCCT 176
lizard     ---CATGGAGCTCGCTTAACCAAAATGTCAATGGAAAAGATTGACTTGGTGGCTCCTCAGCATAGGAACTGTAAGCTATTTGTGAAGCCCGAGGCTGTTTGTGTGAACAGACACC 197
turtle     ---CCTTAAATCAAATTGAG---TTCCAAGCAGCATTTTTACTTTGTCTATCTGTAATTTAGAT-----GATCATAGATGGAATATTTTAAAT--TGTTTTGTGCT-CATACA 168
Coelacanth ---CCTGAAGTAAGTTTGTG---TTGTGTTGTAACCAATTG-CTTAAGGCTGGGAACGATATTCAGT-----GAAAAAATACTTTGTAATATTCGTAATAGGCT-AAAATT 178

```

```

human      TTGCGGGCGGGCCGCTTGCC-----CGGAGCCTGAGAGGATTAATG-----AAAACGTGGCGAGCGAAATGGGGCCAGGGACCTGGAG----CAGGGCGGTGAGGAGAG 271
cat        TTGGTGGCGACGCCGCTAGCC-----CGGAGCCTGAAAGGATTAATG-----AAAATGTGTCCGCGCAAGTGGGG-CTGGGGACCGGGAG----CCGGGAGATTAGG--- 267
armadillo  TTGTAGGCGGGCCGCTTGCC-----CGGAGCCTGAGAGGATTAATG-----AAAAT-TGCCGGGTGATGTGGGGCCAGGGGACCGGGCG----CGGGGAGTTAGGGCA 285
rabbit     TTACGGGCGGGGCGCTTGTCT-----CAGGGCCTGAGAGGATTAATG-----AAAACGAGGCAGCGCAAGTGGGG-CTCGGGGCGCG-----GGCAGGTTAGGGCGA 281
dog        CT-----CGGCGCGCTGGGC-----CGGAGCCCGAGAGGATTAATG-----AAAACGTGTGGGTGAGGTTGGGG-CTGGGGACCGG-----GGCG-----GGGAGA 299
mouse      CG-TAACCGGCGCTGCTGGTT-----TGAGCCTGAGAGGATTAATG-----AAAACGTGGCAGGTGATG-GGGT-CAGGGGACCTGGAG----CGGGGAAGCAGAGGGGA 282
opossum    GGGCAGGGGACGGTTTGGTACT-----GGGAGCCCGAAGTGAATTAATG-----AAAATGATGCAGGGATGGGAGGCTTGGGGCCCGGACTGGCGCTGGGCGCAAGGCAGGG 280
lizard     GTGCCACATACACACATATACAGGTTTTTACTAGATCTGTATGTCTAGTCTCCTAGACTAGACTAGACTAGATATTTAGATATAGATTGCATTCAGCATGGGTAACCTAAGGGTGT 317
turtle     GTGAAATGACATTCACCAAC-----ATTTACTGATAAAAATCT-----AATCCTAGTT--CTAATGTGACTGGGAGAATGACTGGT-----GGGAACATCTTGCAAT- 258
Coelacanth AT-TTATTTATTTGGGTTTTT-----AGGGACTGTGTAAGATAA-----CTTAATCGGTGGTTAATTTAGATAGAAAGACATAGAT-----AAGTTAAAGTGACATT 271

```

```

human      TAGGCAG-----CGGGTGGAGGCTGG-----ACGGGAGGGAGGTTCTAG-----GGAGGCTCTGCCGCGGGCACTGTGAGTCTTGCCGATGATGACGAGACCACT 361
cat        --GGTAG-----CGGGT-AGGCTGG-----AGATGAGGGAGCTTTAG-----GAAGGCTCTGCCGCTGTAACCTGTGAGCCCTGGTTCGATGATGACGTTGACCACT 354
armadillo  AGGGTAGT-----CGGGT-AGGCCGG-----AGAGGCGGAAGGTTTAG-----GGAA-CCTCTGCTGCTG-----TGAGCCCTGGTTCGATGATGACGTTGACCACT 368
rabbit     GAGGGCGG-----GTGGG-AGCCCT-----CTGCCACTGTGACCG-----TGAGCCCTGGCCGATGACGACCCAGCACT 346
dog        CCGGGCGG-----TCGGG-AGGCC-----CTGCC-CTGCAGCG-----CGAGCCCGGGTTCGATGATGACGTTGACCACT 363
mouse      GAGGGACTGA-----GTGGG-AGCCAGGCTTCGGGGAGGGGATAGGTTTAGGG-----TACCACGCGCTCCAGCTG-----CAAGCCCTG-CCAATGATGACATGACCACT 379
opossum    CAGGGAAGGGGCGAGCTTCTGGGCTGTCCCCGTTGGGGAGGGTGGGCGCTGCTGGGGTGGGGGAGGAAGGGGAGCCATAGCGCGCTTTTGGAAACCCTGGTTCGATGATGACGTTGACCACT 400
lizard     TTGGTTGTGGCTACAATTTTAGGATGTAATTTG--TCATACGTTCAAATFACTCCAGTTAG-----AGAAGCATTTCCCTTG--TTGCAAGGACTGGTCTGTGATGAATAGCCCACT 425
turtle     GAAGATGG-----CAGGAAAG-----AAAAAGTGTGGAGCTC---ACTAAGCCCTGGTCTGTGATGATGATTAGCCCACT 322
Coelacanth AAAAATGG-----TAGGAATGATTTT-----TTTTT-----AAATGTGTAGTAATT---TTAAACCAATGGTCTATGATGACTAGCTCACT 346

```

```

human      CCGCAATCTGAGTTCTG---GGAACC--AGGTGATGGAGTATGTTCTGAGAACA--GACTGAGGCCG 421
cat        CCGCAATCTGAGTTCTA---GGAACC--AGGTGATGGAGTGTGTGCTGAGAGCT--GACTGAGACCG 414
armadillo  GCTCAATCTGAGTTCTT---AGAACC--AGGTGATGGAGTGTG-GCTGAGAATG--GACTGAGACCG 427
rabbit     GCACAACCTGAGTCTCTG---CACACC--AGGCTTGGAGTGTGTTCTGAGAACA--GACCAAGGCCG 406
dog        CCGCAATCTGAGTTCTAG---GAACCC--AGGTGATGGAGTGGATGCTGAGAGCG--GACTGAGACCT 424
mouse      CCGCAATCTGAGTTCTG---GGAACC--AGGTGATGGAGTGTG-TGCTGAGAAC---ACTGAGGCCA 435
opossum    CCGCAATCTGAGTTCTG---AGAACC--AGGTGATGGAAATGTGTGCTGAGAGCG--GACTGAGACCA 460
lizard     CCGCAATCTGAGTTCTGT---AGAACC--CAATGATGGATTTTGTGCTTTCGAAGGCATCTGAGACCA 488
turtle     CCGCAATCTGAGTTCTG---AGAACC--TACATGATGAGATTTGTGCTGAGAAGGCAGCTGAGACCA 385
Coelacanth CCGCAATCTGAGTTCTGCGGAGACGTATCATGAGAGAGCAGTGTGCTGACGAAG--AACTGAGACCA 411

```

Figure S5. Alignment of vertebrate mgU12-22/U4-8 RNAs.

```

sense box D'      box C'      box D
** *** ***** *      * :      * : ..      * .** ** : * ****

```