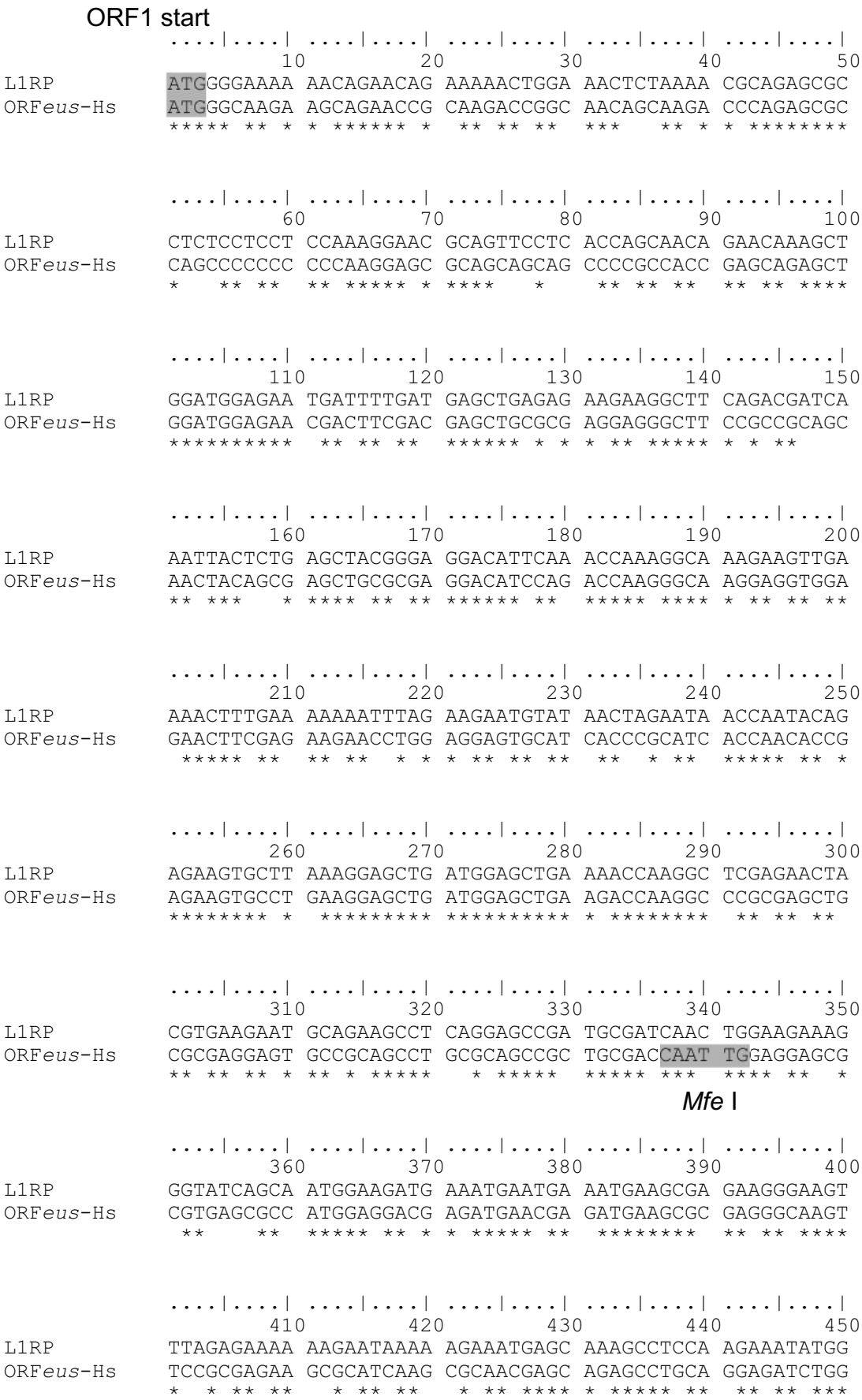


Supplementary Figure 1



```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      460      470      480      490      500
L1RP   GACTATGTGA AAAGACCAAA TCTACGTCTG ATTGGTGTAC CTGAAAGTGA
ORFeus-Hs GACTACGTGA AGCGCCCCAA CCTGCGCCTG ATCGGCGTGC CCGAGAGCGA
      ***** * * * * * ** * * * * ** * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      510      520      530      540      550
L1RP   TGTGGAGAAT GGAACCAAGT TGGAAAAACAC TCTGCAGGAT ATTATCCAGG
ORFeus-Hs CGTGGAGAAC GGCACCAAGC TGGAGAACAC CCTGCAGGAC ATCATCCAGG
      ***** ** ***** ***** ***** ** *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      560      570      580      590      600
L1RP   AGAACTTCCC CAATCTAGCA AGGCAGGCCA ACGTTCAGAT TCAGGAAATA
ORFeus-Hs AGAACTTCCC CAACCTGGCC CGCCAGGCCA ACGTGCAGAT CCAGGAGATC
      ***** *** * * * * * * ***** ***** ***** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      610      620      630      640      650
L1RP   CAGAGAACGC CACAAAGATA CTCCTCGAGA AGAGCAACTC CAAGACACAT
ORFeus-Hs CAGCGCACCC CCCAGCGCTA CAGCAGCCGC CGCGCCACCC CCCGCCACAT
      *** * * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      660      670      680      690      700
L1RP   AATTGTCAGA TTCACCAAAG TTGAAATGAA GGAAAAAATG TTAAGGGCAG
ORFeus-Hs CATCGTGCGC TTCACCAAGG TGGAGATGAA GGAGAAGATG CTGCGCGCCG
      ** * * * ***** * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      710      720      730      740      750
L1RP   CCAGAGAGAA AGGTCGGGTT ACCCTCAAAG GAAAGCCCAT CAGACTAACA
ORFeus-Hs CCCGCGAGAA GGGCCGCGTG ACCCTGAAGG GCAAGCCCAT CCGTCTCACC
      ** * ***** ** * * * ***** * * * * *

```

BsmBI

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      760      770      780      790      800
L1RP   GCGGATCTCT CGGCAGAAAC CCTACAAGCC AGAAGAGAGT GGGGGCCAAT
ORFeus-Hs GCGGACCTGA GCGCCGAGAC CCTGCAGGCC CGCCGCGAGT GGGGCCCCAT
      ** * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      810      820      830      840      850
L1RP   ATTCAACATT CTTAAAGAAA AGAATTTTCA ACCCAGAATT TCATATCCAG
ORFeus-Hs CTTCAACATC CTGAAGGAGA AGAACTTCCA GCCCCGCATC AGCTACCCCG
      ***** ** * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      860      870      880      890      900
L1RP   CCAAACTAAG CTTCATAAGT GAAGGAGAAA TAAAATACTT TATAGACAAG
ORFeus-Hs CCAAGCTGAG CTTCATCAGC GAGGGCGAGA TCAAGTACTT CATCGACAAG
      ***** * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      910      920      930      940      950
L1RP   CAAATGTTGA GAGATTTTGT CACCACCAGG CCTGCCCTAA AAGAGCTCCT
ORFeus-Hs CAGATGCTGC GCGACTTTCGT GACCACCCGC CCCGCCCTGA AGGAGCTGCT
      ** ** * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      960      970      980      990     1000
L1RP   GAAGGAAGCG CTAAACATGG AAAGGAACAA CCGGTACCAG CCGCTGCAAA
ORFeus-Hs GAAGGAGGCC CTGAACATGG AGCGCAACAA CCGTACCAG CCCCTGCAGA
      ***** ** * * * * * * * * * * * * * *

```

ORF1 stop

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1010     1020     1030     1040     1050
L1RP   ATCATGCCAA AATGTAAAGA CCATCAAGAC TAGGAAGAAA CTGCACCAAC
ORFeus-Hs ACCACGCCAA GATGTAAAGA CCATCGAGAC TAGGAAGAAA CTGCATCAAC
      * * * * * * * * * * * * * * * * * * * *

```

ORF2 start

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1060     1070     1080     1090     1100
L1RP   TAATGAGCAA AATCACCAGC TAACATCATA ATGACAGGAT CAACTTCACA
ORFeus-Hs TAATGAGCAA AATCAGGCGC GCCCATCATA ATGACCGGCA GCAACAGCCA
      ***** * * * * * * * * * * * * * *

```

Asc I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1110     1120     1130     1140     1150
L1RP   CATAACAATA TTAAC TTAA ATATAAATGG ACTAAATTCT GCAATTAAAA
ORFeus-Hs CATCACCATC CTGACCCTGA ACATCAACGG CCTGAACAGC GCCATCAAGC
      *** ** * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1160     1170     1180     1190     1200
L1RP   GACACAGACT GGCAAGTTGG ATAAAGAGTC AAGACCCATC AGTGTGCTGT
ORFeus-Hs GCCACCGCCT GGCCAGCTGG ATCAAGAGCC AGGACCCAG CGTGTGCTGC
      * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1210     1220     1230     1240     1250
L1RP   ATTCAGGAAA CCCATCTCAC GTGCAGAGAC ACACATAGGC TCAAAATAAA
ORFeus-Hs ATCCAGGAGA CCCACCTCAC GTGCCGCGAC ACCCACCACC TGAAGATCAA
      ** * * * * * * * * * * * * * * * *

```

Pml I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1260     1270     1280     1290     1300
L1RP   AGGATGGAGG AAGATCTACC AAGCCAATGG AAAACAAAAA AAGGCAGGGG
ORFeus-Hs GGGCTGGCGC AAGATCTACC AGGCCAACGG CAAGCAGAAG AAGGCCGGCG
      * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1310     1320     1330     1340     1350
L1RP   TTGCAATCCT AGTCTCTGAT AAAACAGACT TTAAACCAAC AAAGATCAAA
ORFeus-Hs TGGCCATCCT GGTGAGCGAC AAGACCGACT TCAAGCCCAC CAAGATCAAG
      * * * * * * * * * * * * * * * *

```



```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1810      1820      1830      1840      1850
L1RP    AATCTCACTC AAAGCCGCTC AACTACATGG AAAGTGAACA ACCTGCTCCT
ORFeus-Hs AACCTGACCC AGAGCCGCGAG CACCACCTGG AAGCTGAACA ACCTGCTGCT
      ** ** ** * * ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1860      1870      1880      1890      1900
L1RP    GAATGACTAC TGGGTACATA ACGAAATGAA GGCAGAAAATA AAGATGTTCT
ORFeus-Hs GAACGACTAC TGGGTGCACA ACGAGATGAA GGCCGAGATC AAGATGTTCT
      *** ** ** * * ** ** ** ** ** ** ** ** ** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1910      1920      1930      1940      1950
L1RP    TTGAAACCAA CGAGAACAAA GACACCACAT ACCAGAATCT CTGGGACGCA
ORFeus-Hs TCGAAACCAA CGAGAACAAG GACACCACCT ACCAGAACCT GTGGGACGCC
      * ** ** ** * * ** ** ** * * ** ** ** *

```

***Bst* BI**

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1960      1970      1980      1990      2000
L1RP    TTCAAAGCAG TATGTAGAGG GAAATTTATA GCACTAAATG CCTACAAGAG
ORFeus-Hs TTCAAGGCCG TGTGCCGCGG CAAGTTCATC GCCCTGAACG CCTACAAGCG
      ***** ** * * ** * ** ** ** ** ** ** ** ** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2010      2020      2030      2040      2050
L1RP    AAAGCAGGAA AGATCCAAAA TTGACACCCT AACATCACAA TTAAGAAGAA
ORFeus-Hs CAAGCAGGAG CGCAGCAAGA TCGACACCCT GACCAGCCAG CTGAAGGAGC
      ***** * ** * * ** ** ** ** ** ** * ** *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2060      2070      2080      2090      2100
L1RP    TAGAAAAGCA AGAGCAAAACA CATTCAAAAAG CTAGCAGAAG GCAAGAAATA
ORFeus-Hs TGGAGAAGCA GGAGCAGACC CACAGCAAGG CCAGCCGCCG CCAGGAGATC
      * ** ** ** * ** * ** * * ** * * ** *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2110      2120      2130      2140      2150
L1RP    ACTAAAATCA GAGCAGAACT GAAGGAAATA GAGACACAAA AAACCCCTTCA
ORFeus-Hs ACCAAGATCC GCGCCGAGCT GAAGGAGATC GAGACCCAGA AGACCCCTGCA
      ** ** ** * ** * ** * ** ** ** * ** *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2160      2170      2180      2190      2200
L1RP    AAAAAATCAAT GAATCCAGGA GCTGGTTTTT TGAAAGGATC AACAAAATTG
ORFeus-Hs GAAGATCAAC GAGTCGCGAA GCTGGTTCTT CGAGCGCATC AACAAAGATCG
      ** ** ** * ** * ** * ** * * ** * ** *

```

***Nru* I**

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2210      2220      2230      2240      2250
L1RP    ATAGACCGCT AGCAAGACTA ATAAAGAAAA AAAGAGAGAA GAATCAAATA
ORFeus-Hs ACCGCCCCCT GGCCCGCCTG ATCAAGAAGA AGCGCGAGAA GAACCAGATC
      * * ** ** * ** * ** ** ** * * ** *

```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2260      2270      2280      2290      2300
L1RP   GACACAATAA AAAATGATAA AGGGGATATC ACCACCGATC CCACAGAAAT
ORFeus-Hs GACACCATCA AGAACGACAA GGGCGACATC ACCACCGACC CCACCGAGAT
      ***** ** * * ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2310      2320      2330      2340      2350
L1RP   ACAAACCTACC ATCAGAGAAT ACTACAAACA CCTCTACGCA AATAAACTAG
ORFeus-Hs CCAGACCACC ATCCGCGAGT ACTACAAGCA CCTGTACGCC AACAAGCTGG
      ** ** *** ** * ** * ***** ** ** ***** ** ** ** *
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2360      2370      2380      2390      2400
L1RP   AAAATCTAGA AGAAATGGAT ACATTCTCTG ACACATACAC TCTCCCAAGA
ORFeus-Hs AGAACCTGGA GGAGATGGAC ACCTTCCTGG ACACCTACAC CCTGCCCCGC
      * ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ***** * ***** ***** ** ** *
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2410      2420      2430      2440      2450
L1RP   CTAAACCAGG AAGAAGTTGA ATCTCTGAAT CGACCAATAA CAGGCTCTGA
ORFeus-Hs CTGAACCAGG AGGAGGTGGA GAGCCTGAAC CGCCCCATCA CCGGCAGCGA
      ** ***** * ** ** ** ** ***** ** ** ** * * ** ** *
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2460      2470      2480      2490      2500
L1RP   AATTGTGGCA ATAATCAATA GTTTACCAAC CAAAAAGAGT CCAGGACCAG
ORFeus-Hs GATCGTGGCC ATCATCAACA GCCTGCCCAC CAAGAAGAGC CCCGCCCCCG
      ** ***** ** ***** * * * ** ** ** ***** ** ** ** *
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2510      2520      2530      2540      2550
L1RP   ATGGATTACAG AGCCGAATTC TACCAGAGGT ACAAGGAGGA ACTGGTACCA
ORFeus-Hs ACGGCTTCAC CGCCGAATTC TACCAGCGCT ACAAGGAGGA GCTGGTGCCC
      * ** ***** ***** ** * ***** ***** ***** **
```

EcoR I

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2560      2570      2580      2590      2600
L1RP   TTCCTTCTGA AACTATTCCA ATCAATAGAA AAAGAGGGAA TCCTCCCTAA
ORFeus-Hs TTCCTGCTGA AGCTGTTCCA GAGCATCGAG AAGGAGGGCA TCCTGCCCAA
      ***** ***** * ** ***** ** ** ** ***** * ***** ** **
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2610      2620      2630      2640      2650
L1RP   CTCATTTTAT GAGGCCAGCA TCATTCTGAT ACCAAAGCCG GGCAGAGACA
ORFeus-Hs CAGCTTCTAC GAGGCCAGCA TCATCCTGAT CCCCAAGCCC GGCCGCGACA
      * ** ** ***** ***** ***** ** ***** *** * *****
```

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
      2660      2670      2680      2690      2700
L1RP   CAACCAAAAA AGAGAATTTT AGACCAATAT CCTTGATGAA CATTGATGCA
ORFeus-Hs CCACCAAGAA GGAGAACTTC CGCCCCATCA GCCTGATGAA CATCGACGCC
      * ***** ** ***** ** * ** ** * ***** *** ** **
```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2710      2720      2730      2740      2750
L1RP    AAAATCCTCA ATAAAAATACT GGCAAACCGA ATCCAGCAGC ACATCAAAAA
ORFeus-Hs AAGATCCTGA ACAAGATCCT GGCCAACCGC ATCCAGCAGC ACATCAAGAA
      ** ***** * * ** ** ** *** ***** ***** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2760      2770      2780      2790      2800
L1RP    GCTTATCCAC CATGATCAAG TGGGCTTCAT CCCTGGGATG CAAGGCTGGT
ORFeus-Hs GCTGATCCAC CACGACCAGG TGGGCTTCAT CCCCCGGATG CAGGGCTGGT
      *** ***** ** * ** * ***** ***** ** *****

```

Xma I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2810      2820      2830      2840      2850
L1RP    TCAATATACG CAAATCAATA AATGTAATCC AGCATATAAA CAGAGCCAAA
ORFeus-Hs TCAACATCCG CAAGAGCATC AACGTGATCC AGCACATCAA CCGCGCCAAG
      ***** ** ** **** ** ** ** ***** ** ** * * *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2860      2870      2880      2890      2900
L1RP    GACAAAAACC ACATGATTAT CTCAATAGAT GCAGAAAAAG CCTTTGACAA
ORFeus-Hs GACAAGAACC ACATGATCAT CAGCATCGAC GCCGAGAAGG CCTTCGACAA
      ***** ***** ***** ** * ** ** ** ** ** ***** *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2910      2920      2930      2940      2950
L1RP    AATTCAACAA CCCTTCATGC TAAAAACTCT CAATAAATTA GGTATTGATG
ORFeus-Hs GATCCAGCAG CCCTTCATGC TGAAGACCCT GAACAAGCTG GGCATCGACG
      * * * * ***** * * * * * ** * * * * ** * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2960      2970      2980      2990      3000
L1RP    GGACGTATTT CAAAATAATA AGAGCTATCT ATGACAAACC CACAGCCAAT
ORFeus-Hs GCACCTACTT CAAGATCATC CGCGCCATCT ACGACAAGCC CACCGCCAAC
      * * * * * ***** * * * * * * ***** ** *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3010      3020      3030      3040      3050
L1RP    ATCATACTGA ATGGGCAAAA ACTGGAAGCA TTCCCTTTGA AAACCGGCAC
ORFeus-Hs ATCATCCTGA ACGGCCAGAA GCTGGAGGCC TTCCCCCTGA AGACCGGCAC
      ***** ***** * * * * * ***** ** ***** ** *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3060      3070      3080      3090      3100
L1RP    AAGACAGGGA TGCCCTCTCT CACCGCTCCT ATTCAACATA GTGTTGGAAG
ORFeus-Hs GCGTCAGGGC TGCCCCCTGA GCCCCCTGCT GTTCAACATC GTGCTGGAGG
      * ***** ***** ** ** * * * ***** *****

```

Mlu I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3110      3120      3130      3140      3150
L1RP    TTCTGGCCAG GGCAATCAGG CAGGAGAAGG AAATAAAGGG TATTCAATTA
ORFeus-Hs TGCTGGCCCG CGCCATCCGC CAGGAGAAGG AGATCAAGGG CATCCAGCTG
      * ***** * ** * * * ***** * * * ***** ** * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3160      3170      3180      3190      3200
L1RP      GGAAAAGAGG AAGTCAAATT GTCCCTGTTT GCAGACGACA TGATTGTTTA
ORFeus-Hs GGCAAGGAGG AGGTGAAGCT GAGCCTGTTC GCCGACGACA TGATCGTGTA
      ** ** **** * ** * * * ***** ** ***** **** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3210      3220      3230      3240      3250
L1RP      TCTAGAAAAC CCCATCGTCT CAGCCCCAAA TCTCCTTAAG CTGATAAGCA
ORFeus-Hs CCTGGAGAAC CCCATCGTGA GCGCCCAGAA CCTGCTGAAG CTGATCAGCA
      ** ** *** ***** ***** ** ** ** *** ***** ****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3260      3270      3280      3290      3300
L1RP      ACTTCAGCAA AGTCTCAGGA TACAAAATCA ATGTACAAAA ATCACAAGCA
ORFeus-Hs ACTTCAGCAA GGTGAGCGGC TACAAGATCA ACGTGCAGAA GAGCCAGGCC
      ***** ** ** ***** ** * ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3310      3320      3330      3340      3350
L1RP      TTCTTATACA CCAACAACAG ACAAAACAGAG AGCCAAATCA TGGGTGAACT
ORFeus-Hs TTCTGTACA CCAACAACCG CCAGACCGAG AGCCAGATCA TGGGCGAGCT
      *** * **** ***** * ** ** ** ***** ***** **** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3360      3370      3380      3390      3400
L1RP      CCCATTACACA ATTGCTTCAA AGAGAATAAA ATACCTAGGA ATCCAACCTTA
ORFeus-Hs GCCCTTCACC ATCGCTAGCA AGCGCATCAA GTACCTGGGC ATCCAGCTGA
      ** ***** ** *** * ** * ** ***** ** ***** ** **

```

Nhe I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3410      3420      3430      3440      3450
L1RP      CAAGGGATGT GAAGGACCTC TTCAAGGAGA ACTACAAACC ACTGCTCAAG
ORFeus-Hs CCCGCGACGT GAAGGACCTG TTCAAGGAGA ACTACAAGCC CCTGCTGAAG
      * * ** ** ***** ***** ***** ** ***** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3460      3470      3480      3490      3500
L1RP      GAAATAAAAG AGGAGACAAA CAAATGGAAG AACATTCCAT GCTCATGGGT
ORFeus-Hs GAGATCAAGG AGGAGACCAA CAAGTGGAAG AACATCCCCCT GCAGCTGGGT
      ** ** ** * ***** ** *** ***** ***** ** * ** *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3510      3520      3530      3540      3550
L1RP      AGGAAGAATC AATATCGTGA AAATGGCCAT ACTGCCCAAG GTAATTTACA
ORFeus-Hs GGGCCGCATC AACATCGTGA AGATGGCCAT CCTGCCCAAG GTGATCTACC
      ** * *** ** ***** * ***** ***** ***** ** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3560      3570      3580      3590      3600
L1RP      GATTCAATGC CATCCCCATC AAGCTACCAA TGACTTTCTT CACAGAATTG
ORFeus-Hs GCTTCAACGC CATCCCCATC AAGCTGCCCA TGACCTTCTT CACCGAGCTG
      * ***** ** ***** ***** ** * ***** ***** *** ** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3610      3620      3630      3640      3650
L1RP      GAAAAAACTA CTTTAAAGTT CATATGGAAC CAAAAAAGAG CCCGCATTGC
ORFeus-Hs GAGAAGACCA CCCTGAAGTT CATCTGGAAC CAGAAGCGCG CCCGCATCGC
      ** ** ** * * * * * ** * * * * * ** ** * * * * * **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3660      3670      3680      3690      3700
L1RP      CAAGTCAATC CTAAGCCAAA AGAACAAAAGC TGGAGGCATC ACACTACCTG
ORFeus-Hs CAAGAGCATC CTGAGCCAGA AGAACAAAGGC CGGCGGCATC ACCCTGCCCCG
      ****      *** ** * * * * * * * * * * ** * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3710      3720      3730      3740      3750
L1RP      ACTTCAAAC TACTACAAG GCTACAGTAA CAAAAACAGC ATGGTACTGG
ORFeus-Hs ACTTCAAGCT GTACTACAAG GCCACCGTGA CCAAGACCGC CTGGTACTGG
      ***** ** * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3760      3770      3780      3790      3800
L1RP      TACCAAAACA GAGATATAGA TCAATGGAAC AGAACAGAGC CCTCAGAAAT
ORFeus-Hs TACCAGAACC GCGATATCGA CCAGTGGAAC CGCACCGAGC CCAGCGAGAT
      ***** ** * * * * * * * * * * * * * * * *

```

EcoR V

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3810      3820      3830      3840      3850
L1RP      AATGCCGCAT ATCTACAACT ATCTGATCTT TGACAAACCT GAGAAAAACA
ORFeus-Hs CATGCCCCAC ATCTACAACT ACCTGATCTT CGACAAGCCC GAGAAGAACA
      ***** ** * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3860      3870      3880      3890      3900
L1RP      AGCAATGGGG AAAGGATTCC CTATTTAATA AATGGTGCTG GGAAAACTGG
ORFeus-Hs AGCAGTGGGG CAAGGACAGC CTGTTCAACA AGTGGTGCTG GGAGAACTGG
      **** * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3910      3920      3930      3940      3950
L1RP      CTAGCCATAT GTAGAAAGCT GAAACTGGAT CCCTTCCTTA CACCTTATAC
ORFeus-Hs CTGGCCATCT GCCGCAAGCT GAAGCTGGAT CCCTTCCTGA CCCCTACAC
      ** * * * * * * * * * * * * * * * * * * * *

```

BamH I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      3960      3970      3980      3990      4000
L1RP      AAAAAATCAAT TCAAGATGGA TTAAAGATTT AAACGTTAAA CCTAAACCA
ORFeus-Hs CAAGATCAAC AGCCGCTGGA TCAAGGACCT GAACGTGAAG CCCAAGACCA
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4010      4020      4030      4040      4050
L1RP      TAAAAACCTT AGAAGAAAAC CTAGGCATTA CCATTCAGGA CATAGGCGTG
ORFeus-Hs TCAAGACCTT GGAGGAGAAC CTGGGCATCA CCATCCAGGA CATCGGCGTG
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4060      4070      4080      4090      4100
L1RP   GGCAAGGACT TCATGTCCAA AACACCAAAA GCAATGGCAA CAAAAGACAA
ORFeus-Hs GGCAAGGACT TCATGAGCAA GACCCCCAAG GCCATGGCCA CCAAGGACAA
      ***** ***** *** ** * * * ** ***** * * * * *****

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4110      4120      4130      4140      4150
L1RP   AATTGACAAA TGGGATCTAA TTAAACTAAA GAGCTTCTGC ACAGCAAAAAG
ORFeus-Hs GATCGACAAG TGGGACCTGA TCAAGCTGAA GAGCTTCTGC ACCGCCAAGG
      ** ***** ***** * * * * * * * * * * ***** ** * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4160      4170      4180      4190      4200
L1RP   AACTTACCAT CAGAGTGAAC AGGCAACCTA CAACATGGGA GAAAATTTTT
ORFeus-Hs AGACCACCAT CCGCGTGAAC CGCCAGCCCA CCACCTGGGA GAAGATCTTC
      * * * ***** * * * * * * * * * * * * * * ***** *** * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4210      4220      4230      4240      4250
L1RP   GCAACCTACT CATCTGACAA AGGGCTAATA TCCAGAATCT ACAATGAACT
ORFeus-Hs GCCACCTACA GCAGCGACAA GGGCCTGATC AGCCGCATCT ACAACGAGCT
      ** ***** ***** ** * * * * * * * * * * ***** * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4260      4270      4280      4290      4300
L1RP   CAAACAAATT TACAAGAAAA AAACAAACAA CCCCATCAAA AAGTGGGCGA
ORFeus-Hs GAAGCAGATC TACAAGAAGA AGACCAACAA CCCCATCAAG AAGTGGGCCA
      ** * * * ***** * * * * * * * * * * ***** ***** *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4310      4320      4330      4340      4350
L1RP   AGGACATGAA CAGACACTTC TCAAAAAGAAG ACATTTATGC AGCCAAAAAA
ORFeus-Hs AGGACATGAA CCGCCACTTC AGCAAGGAGG ACATCTACGC CGCCAAGAAG
      ***** * * * ***** ** * * * * * * * * * * ***** **

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4360      4370      4380      4390      4400
L1RP   CACATGAAGA AATGCTCATC ATCACTGGCC ATCAGAGAAA TGCAAATCAA
ORFeus-Hs CATATGAAGA AGTGCAGCAG CAGCCTGGCC ATCCGCGAGA TGCAGATCAA
      ** ***** * * * ***** ***** *** * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4410      4420      4430      4440      4450
L1RP   AACCACATATG AGATATCATC TCACACCAGT TAGAATGGCA ATCATTAATA
ORFeus-Hs GACCACCATG CGCTACCACC TGACCCCCGT GCGCATGGCC ATCATCAAGA
      ***** *** * * * * * * * * * * * * * * ***** ***** *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4460      4470      4480      4490      4500
L1RP   AGTCAGGAAA CAACAGGTGC TGGAGAGGAT GCGGAGAAAT AGGAACACTT
ORFeus-Hs AGAGCGGCAA CAACCGCTGC TGGCGCGGCT GCGGCGAGAT CGGCACCCTG
      ** * * * * * * * * * * * * * * * * * * * * ***** * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4510      4520      4530      4540      4550
L1RP      TTACACTGTT GGTGGGACTG TAAACTAGTT CAACCATTTG GGAAGTCAGT
ORFeus-Hs CTGCACTGCT GGTGGGACTG CAAGCTGGTG CAGCCCCTGT GGAAGAGCGT
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4560      4570      4580      4590      4600
L1RP      GTGGCGATTG CTCAGGGATC TAGAACTAGA AATACCATTT GACCCAGCCA
ORFeus-Hs GTGGCGCTTC CTGCGCGACC TGGAGCTGGA GATCCCCTTC GACCCGCCA
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4610      4620      4630      4640      4650
L1RP      TCCCATTACT GGGTATATAC CCAAATGAGT ATAAATCATG CTGCTATAAA
ORFeus-Hs TCCCCCTGCT GGCATCTAC CCCAACGAGT ACAAGAGCTG CTGCTACAAG
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4660      4670      4680      4690      4700
L1RP      GACACATGCA CACGTATGTT TATTGCGGCA CTATTCACAA TAGCAAAGAC
ORFeus-Hs GACACCTGCA CCCGCATGTT CATCGCCGCC CTGTTACCA TCGCCAAGAC
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4710      4720      4730      4740      4750
L1RP      TTGGAACCAA CCCAAATGTC CAACAATGAT AGACTGGATT AAGAAAATGT
ORFeus-Hs CTGGAACCAG CCCAAGTGCC CCACCATGAT CGATTGGATC AAGAAGATGT
      * * * * * * * * * * * * * * * * * * * *

```

Cla I

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4760      4770      4780      4790      4800
L1RP      GGCACATATA CACCATGGAA TACTATGCAG CCATAAAAAA TGATGAGTTC
ORFeus-Hs GGCACATCTA CACCATGGAG TACTACGCCG CCATCAAGAA CGACGAGTTC
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4810      4820      4830      4840      4850
L1RP      ATATCCTTTG TAGGGACATG GATGAAATTG GAAACCATCA TTCTCAGTAA
ORFeus-Hs ATCAGCTTCG TGGGCACCTG GATGAAGCTG GAGACCATCA TCCTGAGCAA
      * * * * * * * * * * * * * * * * * * * *

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      4860      4870      4880      4890      4900
L1RP      ACTATCGCAA GAACAAAAAA CCAAACACCG CATATTCTCA CTCATAGGTG
ORFeus-Hs GCTGAGCCAG GAGCAGAAGA CCAAGCACCG CATCTTCAGC CTGATCGGCG
      * * * * * * * * * * * * * * * * * * * *

```

ORF2 stop

....|...

```

L1RP      GGAATTGA
ORFeus-Hs GCAACTGA
      * * * *

```

Supplementary Figure 2

CMV start	
	
	1020304050
CMV	GTTGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT TACGGGGTCA
CMV+5' UTR	GTTGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT TACGGGGTCA
5' UTR	-----
	
	60708090100
CMV	TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
CMV+5' UTR	TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
5' UTR	-----
	
	110120130140150
CMV	TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA
CMV+5' UTR	TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA
5' UTR	-----
	
	160170180190200
CMV	TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA
CMV+5' UTR	TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA
5' UTR	-----
	
	210220230240250
CMV	TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGTGTA
CMV+5' UTR	TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGTGTA
5' UTR	-----
	
	260270280290300
CMV	TCATATGCCA AGTCCGCCCC CTATTGACGT CAATGACGGT AAATGGCCCCG
CMV+5' UTR	TCATATGCCA AGTCCGCCCC CTATTGACGT CAATGACGGT AAATGGCCCCG
5' UTR	-----
	
	310320330340350
CMV	CCTGGCATTG TGCCCACTAC ATGACCTTAC GGGACTTTCC TACTTGGCAG
CMV+5' UTR	CCTGGCATTG TGCCCACTAC ATGACCTTAC GGGACTTTCC TACTTGGCAG
5' UTR	-----
	
	360370380390400
CMV	TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
CMV+5' UTR	TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
5' UTR	-----
	
	410420430440450
CMV	GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC
CMV+5' UTR	GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC
5' UTR	-----

	
	460 470 480 490 500
CMV	TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG
CMV+5' UTR	TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG
5' UTR	-----

	
	510 520 530 540 550
CMV	GACTTTCCAA AATGTCGTAA TAACCCCGCC CCGTTGACGC AAATGGGCGG
CMV+5' UTR	GACTTTCCAA AATGTCGTAA TAACCCCGCC CCGTTGACGC AAATGGGCGG
5' UTR	-----

	
	560 570 580 590 600
CMV	TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT TTAGTGAACC
CMV+5' UTR	TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT TTAGTGAACC
5' UTR	-----

	
	610 620 630 640 650
CMV	GTCAGATCTC TAGAAGCTGG GTACCAGCTG CTAGCAAGCT TGCTAGCGGC
CMV+5' UTR	GTCAGATCTC TAGAAGCTGG GTACCAGCTG CTAGCAAGCT TGCTAGCGGC
5' UTR	-----

Kozak sequence

	
	660 670 680 690 700
CMV	CGCGTTTAAA CTTAATTAAG CCACC ORF1 start.....
CMV+5' UTR	CGCTCTAGCC CTGGAATGTG TATTAAGACT GTAAGGTGGG GGGGAGGAGC
5' UTR	---TCTAGCC CTGGAATGTG TATTAAGACT GTAAGGTGGG GGGGAGGAGC

5' UTR start

	
	710 720 730 740 750
CMV	CAAGATGGCC GAATAGGAAC AGCTCCGGTC TACAGCTCCC AGCGTGAGCG
CMV+5' UTR	CAAGATGGCC GAATAGGAAC AGCTCCGGTC TACAGCTCCC AGCGTGAGCG
5' UTR	-----

	
	760 770 780 790 800
CMV	ACGCAGAAGA CGGTGATTTC TGCATTTCCA TCTGAGGTAC CGGGTTTCATC
CMV+5' UTR	ACGCAGAAGA CGGTGATTTC TGCATTTCCA TCTGAGGTAC CGGGTTTCATC
5' UTR	-----

	
	810 820 830 840 850
CMV	TCACTAGGGA GTGCCAGACA GTGGGCGCAG GCCAGTGTGT GTGCGCACCG
CMV+5' UTR	TCACTAGGGA GTGCCAGACA GTGGGCGCAG GCCAGTGTGT GTGCGCACCG
5' UTR	-----

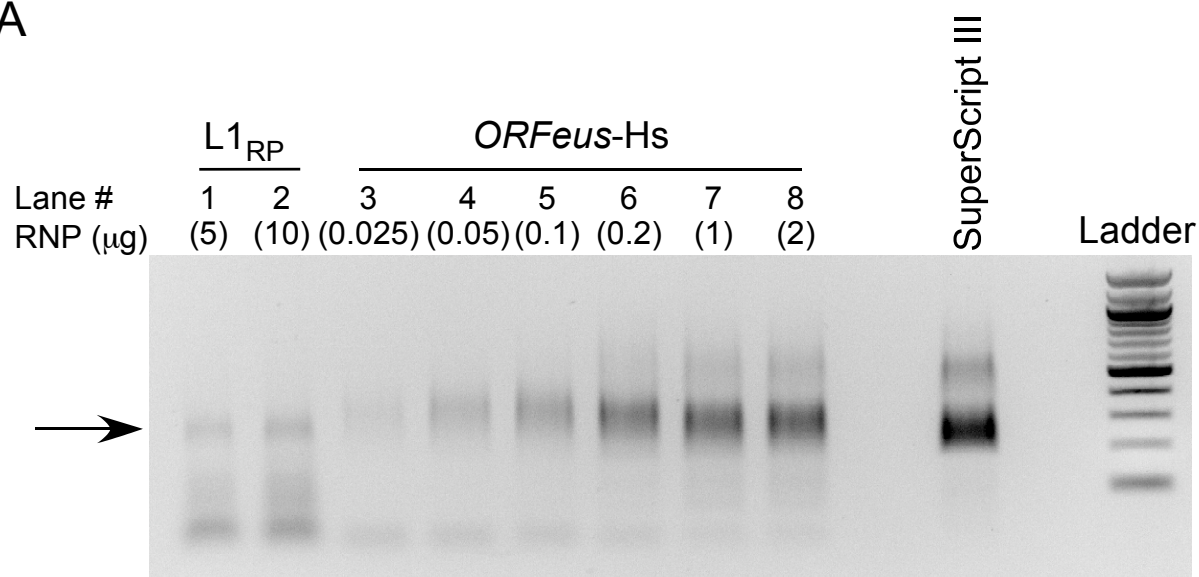
	
	860 870 880 890 900
CMV	TGCGCGAGCC GAAGCAGGGC GAGGCATTGC CTCACCTGGG AAGCGCAAGG
CMV+5' UTR	TGCGCGAGCC GAAGCAGGGC GAGGCATTGC CTCACCTGGG AAGCGCAAGG
5' UTR	-----

	
	910 920 930 940 950
CMV	
CMV+5' UTR	GGTCAGGGAG TTCCCTTTCC GAGTCAAAGA AAGGGGTGAC GGACGCACCT
5' UTR	GGTCAGGGAG TTCCCTTTCC GAGTCAAAGA AAGGGGTGAC GGACGCACCT
	
	960 970 980 990 1000
CMV	
CMV+5' UTR	GGAAAATCGG GTCACTCCCA CCCGAATATT GCGCTTTTCA GACCGGCTTA
5' UTR	GGAAAATCGG GTCACTCCCA CCCGAATATT GCGCTTTTCA GACCGGCTTA
	
	1010 1020 1030 1040 1050
CMV	
CMV+5' UTR	AGAAACGGCG CACCACGAGA CTATATCCCG CACCTGGCTC GGAGGGTCCT
5' UTR	AGAAACGGCG CACCACGAGA CTATATCCCG CACCTGGCTC GGAGGGTCCT
	
	1060 1070 1080 1090 1100
CMV	
CMV+5' UTR	ACGCCACGG AATCTCGCTG ATTGCTAGCA CAGCAGTCTG AGATCAAAC
5' UTR	ACGCCACGG AATCTCGCTG ATTGCTAGCA CAGCAGTCTG AGATCAAAC
	
	1110 1120 1130 1140 1150
CMV	
CMV+5' UTR	GCAAGGCGGC AACGAGGCTG GGGGAGGGGC GCCC GCCATT GCCCAGGCTT
5' UTR	GCAAGGCGGC AACGAGGCTG GGGGAGGGGC GCCC GCCATT GCCCAGGCTT
	
	1160 1170 1180 1190 1200
CMV	
CMV+5' UTR	GCTTAGGTAA ACAAAGCAGC AGGGAAGCTC GAACTGGGTG GAGCCCACCA
5' UTR	GCTTAGGTAA ACAAAGCAGC AGGGAAGCTC GAACTGGGTG GAGCCCACCA
	
	1210 1220 1230 1240 1250
CMV	
CMV+5' UTR	CAGCTCAAGG AGGCCTGCCT GCCTCTGTAG GCTCCACCTC TGGGGGCAGG
5' UTR	CAGCTCAAGG AGGCCTGCCT GCCTCTGTAG GCTCCACCTC TGGGGGCAGG
	
	1260 1270 1280 1290 1300
CMV	
CMV+5' UTR	GCACAGACAA ACAAAAAGAC AGCAGTAACC TCTGCAGACT TAAGTGTCCC
5' UTR	GCACAGACAA ACAAAAAGAC AGCAGTAACC TCTGCAGACT TAAGTGTCCC
	
	1310 1320 1330 1340 1350
CMV	
CMV+5' UTR	TGTCTGACAG CTTTGAAGAG AGCAGTGGTT CTCCCAGCAC GCAGCTGGAG
5' UTR	TGTCTGACAG CTTTGAAGAG AGCAGTGGTT CTCCCAGCAC GCAGCTGGAG

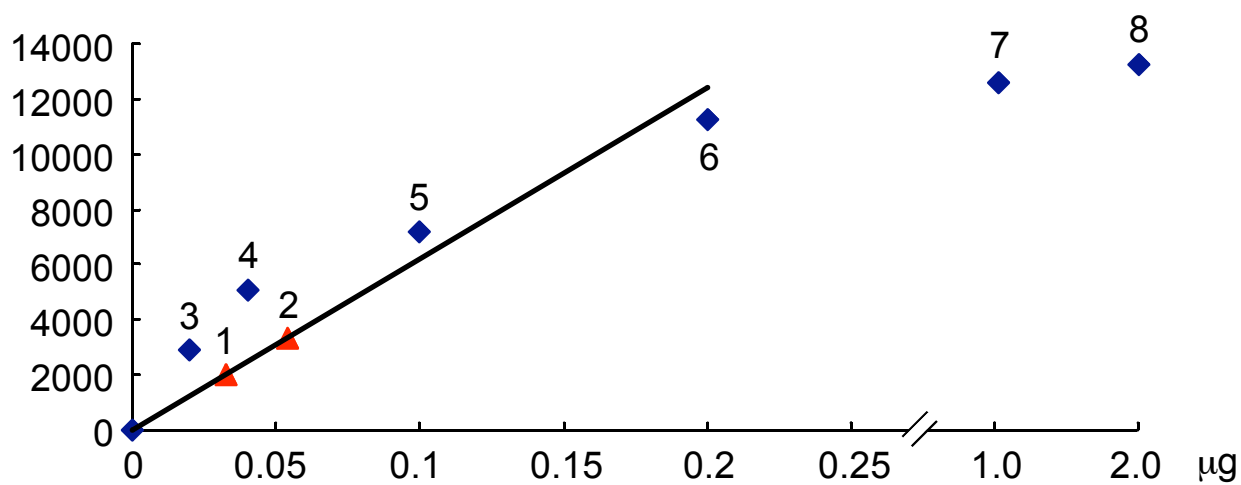
		
	1360 1370 1380 1390 1400	
CMV		
CMV+5' UTR	ATCTGAGAAC GGGCAGACTG CCTCCTCAAG TGGGTCCCTG ACCCCTGACC	
5' UTR	ATCTGAGAAC GGGCAGACTG CCTCCTCAAG TGGGTCCCTG ACCCCTGACC	
		
	1410 1420 1430 1440 1450	
CMV		
CMV+5' UTR	CCCGAGCAGC CTAACTGGGA GGCACCCCCC AGCAGGGGCA CACTGACACC	
5' UTR	CCCGAGCAGC CTAACTGGGA GGCACCCCCC AGCAGGGGCA CACTGACACC	
		
	1460 1470 1480 1490 1500	
CMV		
CMV+5' UTR	TCACACGGCA GGGTATTCCA ACAGACCTGC AGCTGAGGGT CCTGTCTGTT	
5' UTR	TCACACGGCA GGGTATTCCA ACAGACCTGC AGCTGAGGGT CCTGTCTGTT	
		
	1510 1520 1530 1540 1550	
CMV		
CMV+5' UTR	AGAAGGAAAA CTAACAACCA GAAAGGACAT CTACACCGAA AATCCATCTG	
5' UTR	AGAAGGAAAA CTAACAACCA GAAAGGACAT CTACACCGAA AATCCATCTG	
		
	1560 1570 1580 1590 	
CMV		
CMV+5' UTR	TACATCACCA TCATCAAAGA CAAAAAGTAG ATAAAAACCAC AAAG <u>ORF1 start</u>	
5' UTR	TACATCACCA TCATCAAAGA CAAAAAGTAG ATAAAAACCAC AAAG <u>ORF1 start</u>	

Supplementary Figure 3

A



B



Supplementary Figure 4

