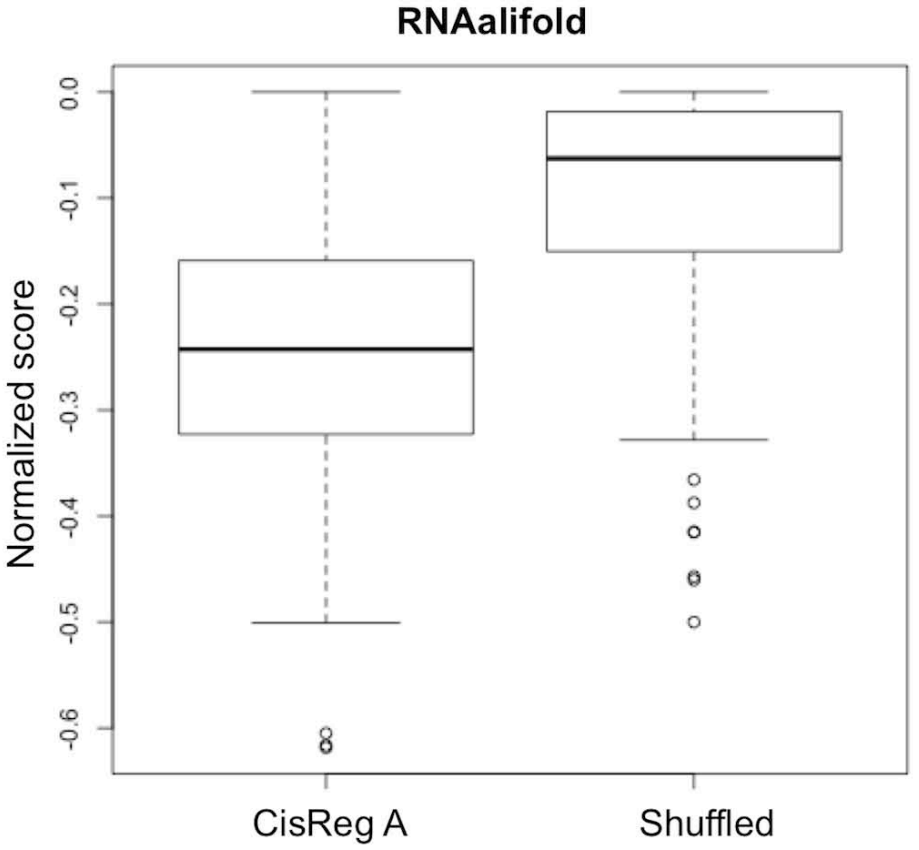
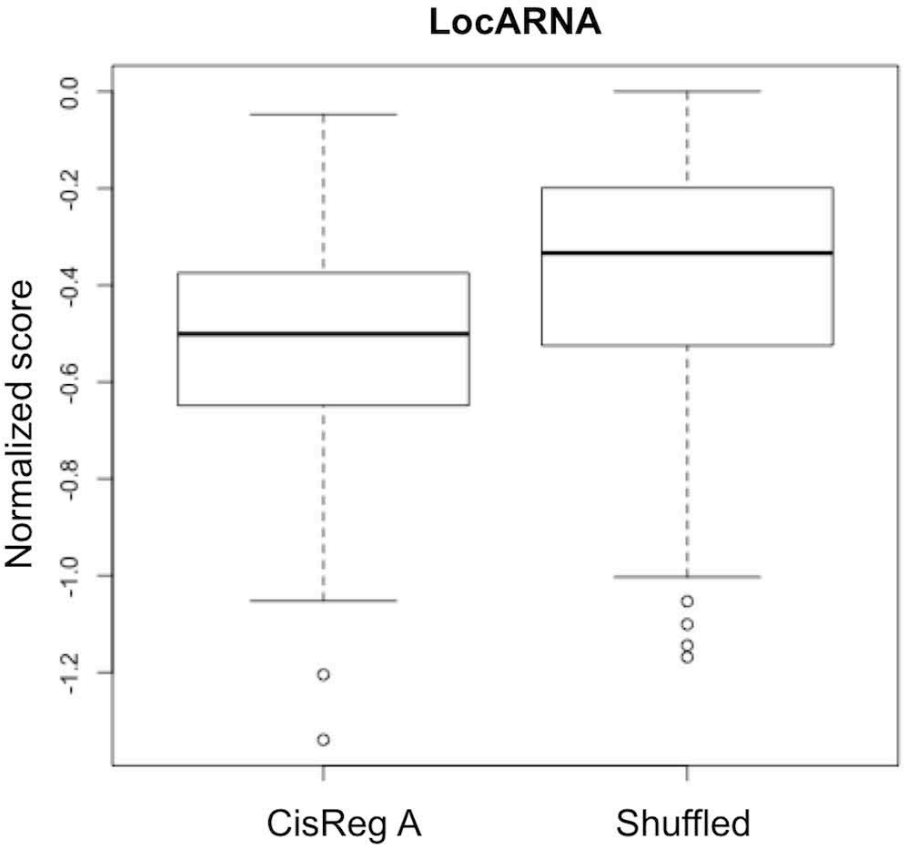
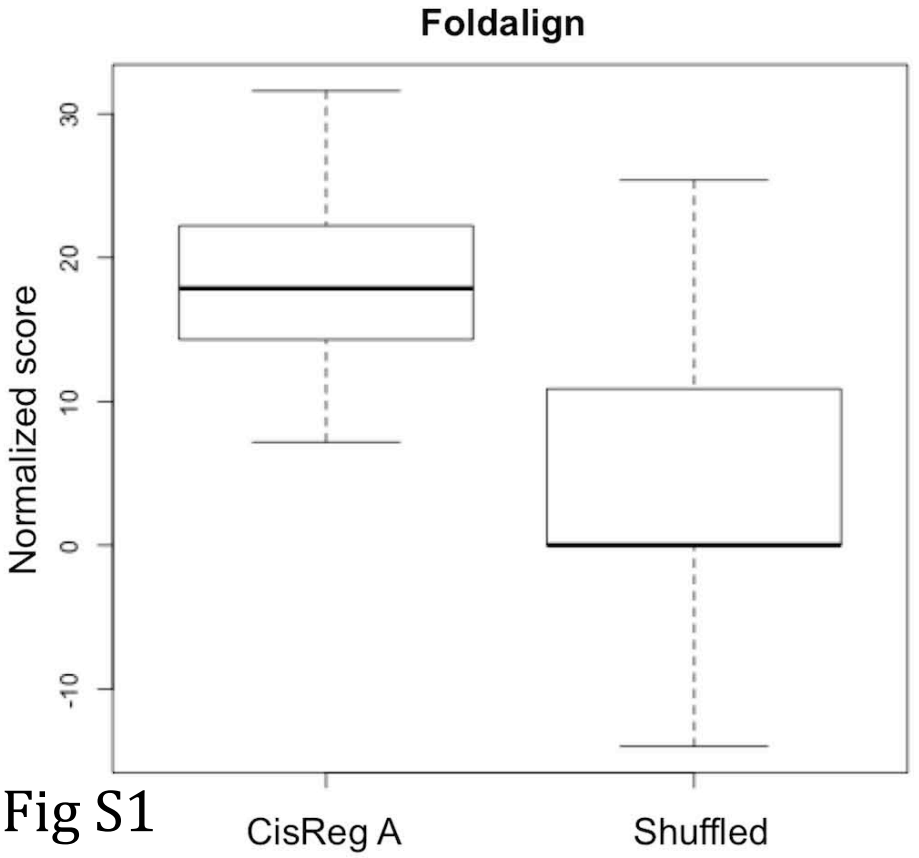

RNAalifold
Normalized score
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
CisReg A
Shuffled


LocARNA
Normalized score
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
CisReg A
Shuffled


Foldalign
Normalized score
30
20
10
0
-10
CisReg A
Shuffled

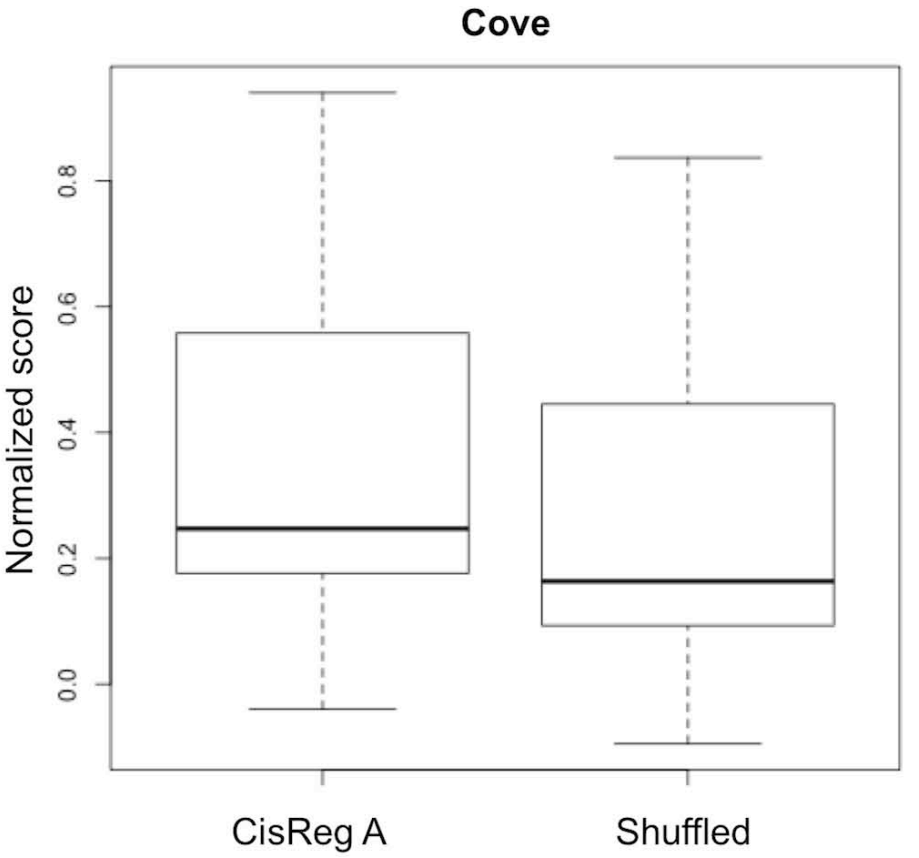

Cove
Normalized score
0.8
0.6
0.4
0.2
0.0
CisReg A
Shuffled

Fig S1

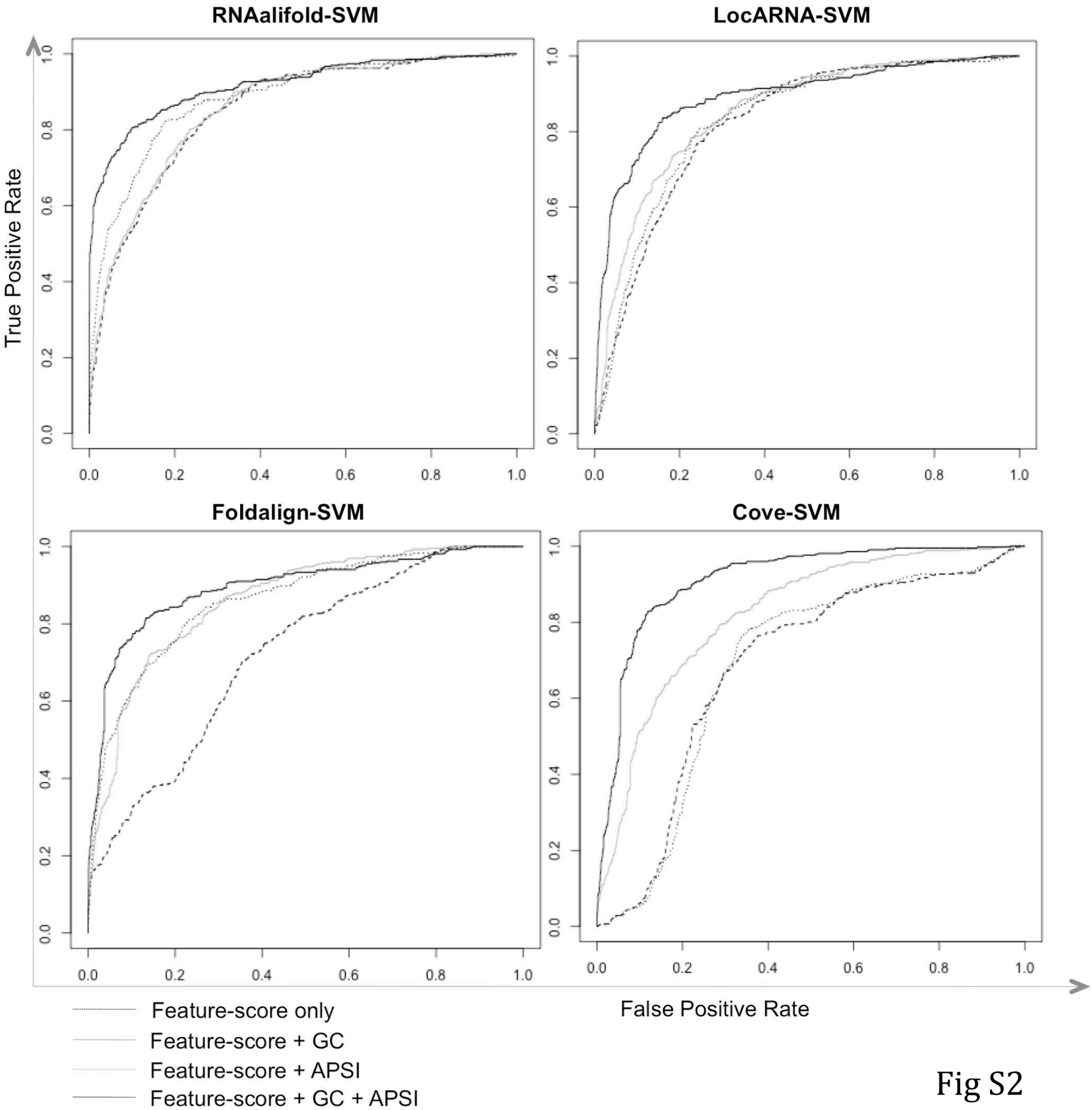
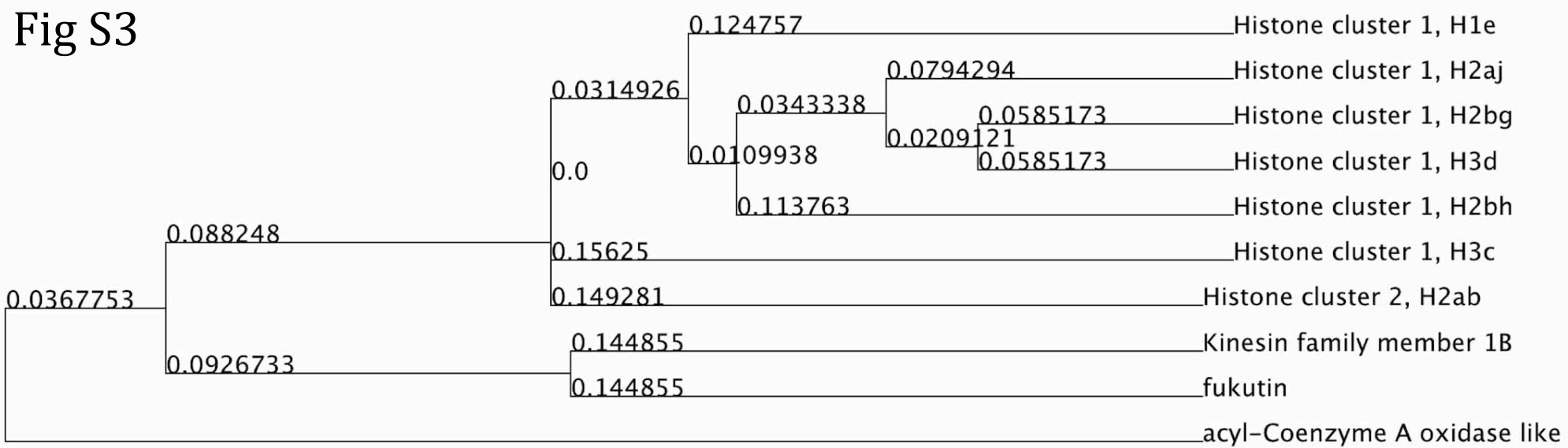

RNAalifold-SVM
LocARNA-SVM
Foldalign-SVM
Cove-SVM
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Feature-score only
Feature-score + GC
Feature-score + APSI
Feature-score + GC + APSI

Fig S2

Fig S3


0.124757
Histone cluster 1, H1e
0.0794294
Histone cluster 1, H2aj
0.0314926
0.0343338
0.0585173
Histone cluster 1, H2bg
0.0209121
0.0109938
0.0585173
Histone cluster 1, H3d
0.0
0.113763
Histone cluster 1, H2bh
0.088248
0.15625
Histone cluster 1, H3c
0.0367753
0.149281
Histone cluster 2, H2ab
0.144855
Kinesin family member 1B
0.0926733
0.144855
fukutin
acyl-Coenzyme A oxidase like

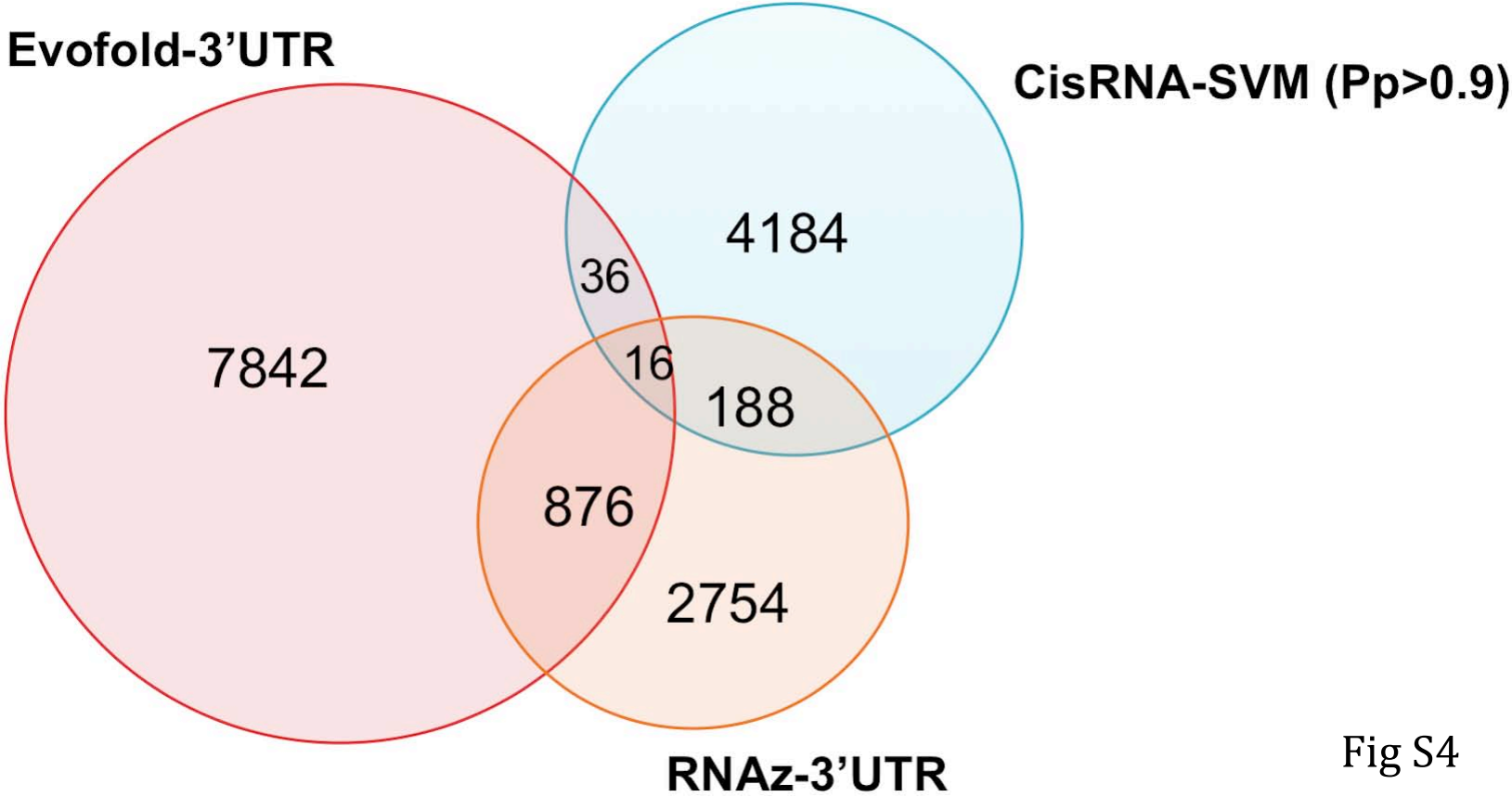

Evofold-3'UTR
CisRNA-SVM (Pp>0.9)
RNAz-3'UTR
7842
4184
2754
36
16
188
876

Fig S4

Clusters:

Cut at a point where there are most clusters with >0.5 average pairwise identity:

Top ranked clusters by RNAalifold minimum free energy:

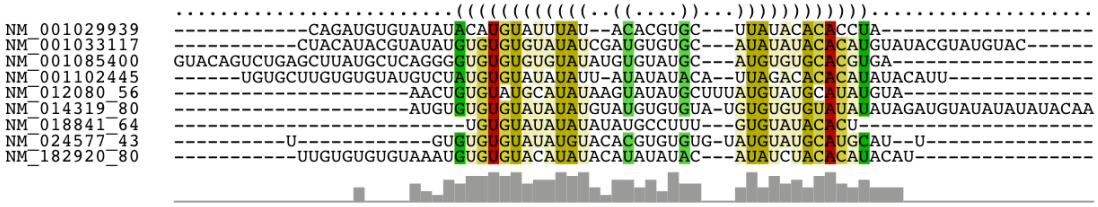
1. NM000767: cytochrome p450, subfamily B (CYP2B6), drug metabolic process, oxidation-reduction process, steroid metabolic process
   NM017964: solute carrier family 30 (zinc transporter) (SLC30A6), cation transmembrane transporter activity
   NM080876: dual specificity phosphatase 19 (DUSP19), regulation of MAPK, JNK cascades, JUN kinase phosphatase activity, tyrosine/threonine phosphatase activity

```
                 .....((..((((((((((((.((....)))).)))))))..)))).)).......
NM_000767_24     GCCCAGGCUGGAGUGCAGUGGCGUGAUCUCGGCUCACUGCAACCUCCACCCCCGGGG
NM_017964_64     ---CAGGCUGGAAUGCAGUGGCAUGAUCUCAGCUCACUGCAACCUCUGCCU-----C
NM_080876_26     ----AGGCUGGAGUGCAGUGGCGUGAUCUUGGCUCACUGCAACCUCUGCCU------
```

2. NM000112: solute carrier family 26 (SLC26A2), sulfate transport, membrane
   NM001235: serpin peptidase inhibitor, clade H (heat shock protein 47)(SERPINH1), negative regulation of endopeptidase activity, endoplasmic reticulum, collagen binding
   NM019035: protocadherin 18 (PCDH18), cell adhesion, membrane, calcium ion binding
   NM145034: torsin A interacting protein 2 (TOR1AIP2), endoplasmic reticulum membrane

```
                 ................(((((((.......((((((.............)))))))......)))....)))))).
NM_000112_0      U---------------UCAAGUGUCC-AACAUUUCCCAGUUCCACA---GUGGGAAAUUUUGCACA----CUUGAA
NM_001235_64     UGAACUUUUUGUUUUGUUUCUUCC-------UUUUUUAGUUCUUCAAAGAUAGGGAG------GGA----AGGGGG
NM_019035_0_     A---------------UUUUU----------GUUUCCAUGUAU------AUGGAAAU-------------AGGGAA
NM_145034_0_     A---------------AAGUUUGUUUAAAGACCUUCCAUUUAU------AUGGAAGGAGGUUGAUAAAAUAGCUUU
```

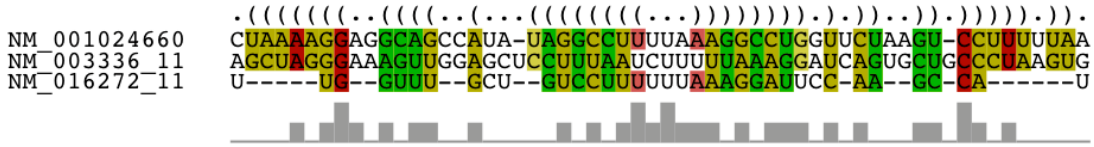
3. NM001029939: zinc finger and AT hook domain containing (ZFAT), zinc ion binding, intracellular
   NM001033117: SLIT-ROBO Rho GTPase activating protein 3 (SRGAP3), GTPase activator activity, signal transduction
   NM001085400: RELT-like 1 (RELL1), microtubule cytoskeleton
   NM001102445: regulator of G-protein signalling 4 (RGS4), GTPase activator activity
   NM012080: haloacid dehalogenase-like hydrolase domain containing 1 (HDHD1), hydrolase activity
   NM014319: LEM domain containing 3 (LEMD3), negative regulation of activin receptor, BMP, transforming growth factor beta receptor signalling pathways
   NM018841: G protein, gamma 12 (GNG12), cellular response to glucagon stimulus, cerebral cortex development, signal transduction
   NM024577: SH3 domain and tetratricopeptide repeats 2 (SH3TC2), binding
   NM182920: ADAM metallopeptidase with thrombospondin type 1 motif (ADAMTS9), zinc ion binding



4. NM001024660: kalirin, RhoGEF kinase (KALRN), small GTPase mediated signal transduction, nerve growth factor recpetor signalling pathway, apoptotic process

NM003336: ubiquitin-conjugaing enzyme E2A (UBE2A), DNA repair, chromatin modification

NM016272: transducer of ERBB2, (TOB2), female gamete generation, negative regulation of cell proliferation

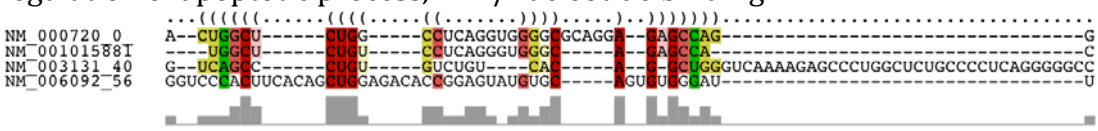


5. NM000720: calcium channel, votage-dependent L type, (CACNA1D), ion transport, membrane

NM001015881: TSC22 domain family member 3(TSC22D3), sequence-specific DNA binding transcription factor activity

NM003131: serum responsive factor (c-fos serum response element-binding transcription factor) (SRF), actin filament organization, developmental growth, long-term memory, platelet formation

NM006092: nucleotide-binding oligomerization domain containing 1 (NOD1), regulation of apoptotic process, ATP/nucleotide binding



6. NM002565: pyrimidinergic receptor P2Y, G-protein coupled (P2RY4), plasma membrane, activation of phospholipase C activity

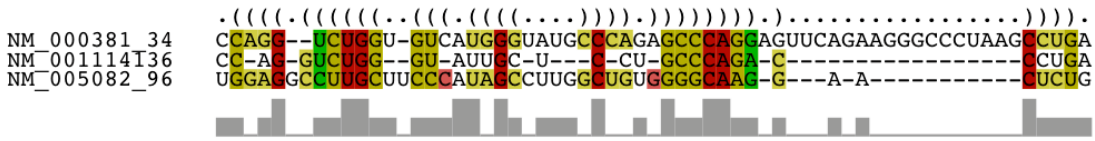
NM014840: NAUK family, SNF1-like kinase (NUAK1), regulation of cell adhesion, proliferation, metal ion binding, p53 binding

NM020336: Ral GTPase activating protein, (RALGAPB), GTPase activator activity

7. NM000381: midline 1 (Opitz/BBB syndrome) (MID1), zinc ion binding

NM001114136: erythrocyte membrane protein band 4.9 (EPB49), actin filament bundle assembly

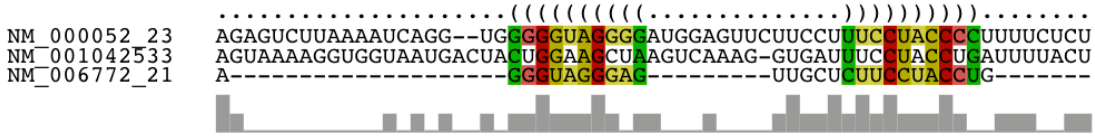
NM005082: tripartite motif containing 25 (TRIM25), innate immune response, regulation of transcription



8. NM000052: ATPase, Cu++ transporting (ATP7A), membrane, ATP metabolic process

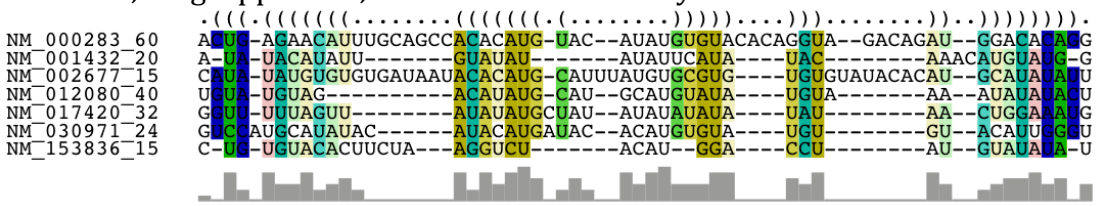
NM001042533: MYC induced nuclear antigen (MINA), ribosome biogenesis

NM006772: synaptic Ras GTPase activating protein 1 (SYNGAP1), dendrite development



Other large clusters (>5 sequences):

Cluster3149:
NM000283: phosphodiesterase 6B, cGMP-specific (PDE6B), detection of light stimulus, retina development in camera-type eye
NM001432: epiregulin (EREG), anatomical structure morphogenesis, cell differentiation, regulation of cell proliferation, transcription, ovulation
NM002677: peripheral myelin protein 2 (PMP2), fatty acid, cholesterol binding, transporter
NM012080: haloacid dehalogenase-like hydrolase domain containing 1 (HDHD1), hydrolase activity
NM017420: SIX homeobox 4 (SIX4), inner ear morphogenesis, regulation of transcription, muscle tissue development
NM030971: sideroflexin 3 (SFXN3), cation transmembrane transporter, mitochondrial membrane
NM153836: cellular repressor of E1A-simulated genes 2 (CREG2), endoplasmic reticulum, Golgi apparatus, oxidoreductase activity



Cluster523:
NM001001709: chromosome 9 open reading frame 170
NM001008239: chromosome 18 open reading frame 25
NM001012418: myosin light chain kinase family (MYLK4), ATP/nucleotide binding
NM001080392: KIAA1147
NM152345: ankyrin repeat domain 13B (ANKRD13B)
NM194249: dead end homolog 1 (zebrafish) (DND1), germ cell development, mRNA/AU-rich element binding

```
                 ((((((((((((((((.............................)))))).)).)).....))))))
NM_001001709     GGUUAGGGGGACU-------AG-------------------CAAGUACC-CU----CUGGCG
NM_001008239     AGGUGUUCUGGGCAUCUGAAGGUAUUAAGAGUUCAGUGGG-AAAAAUCUCUCUCCCUCGCCU
NM_001012418     CGGUGGUGGUGAUGUUGGGAAG--------------------AUGUUUCCCUG----CCAUCU
NM_001080392     GGGGAGAGGAGGAGUU--AAGG--------------------GCUCCGUU-UA----CUCCCA
NM_152345_72     UGGGGGGC-------------AG--------------------CUGA-----GC----UCCCCG
NM_194249_40     UUGUGGGGAAGGGCAGCUGAGCCA-AAGGGGUCAGAAAUUCUGCCCUUUGCCUCCACCACAU
```

Cluster4059: (same as cluster6 by average pairwise identity, see below)

By average pairwise identity:

1. Same as above cluster 1

2. NM003423: zinc finger protein 43 (ZNF43), regulation of transcription
   NM003430: zinc finger protein 91 (ZNF91), regulation of transcription
   NM138330: zinc finger protein 675 (ZNF675), born resorption, cytokine-mediated signalling pathway

```
                 .((((((((((((...((((((......))))))...((((....).)))..)))))))))))).
NM_003423_0      G---UCACACUUGAUUGUAGGUAAGAUAAUUUACAUUGGAGUAAAC-UUCUACAAGUGUGA---A
NM_003430_40     U--UUCACACCUUACUACAGAUAAGAUAACUCAUAUUGGAGAGAAAUCU-UACAAGUGUGAA--U
NM_138330_0_     GUUUUCACACUUCAUUGUAUGUAAGAUAAUUCAUACUGGAGAAAAC----UACAAGUGUGAAGAA
```

3. NM020394: zinc finger protein 695 (ZNF695), regulation of transcription
   NM173531: zinc finger protein 100 (ZNF100), regulation of transcription
   NM178558: zinc finger protein 680 (ZNF680), regulation of transcription

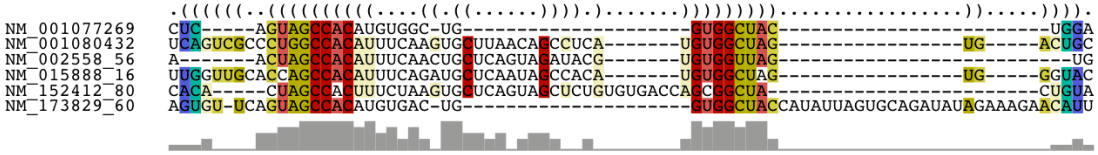
```
                 ............(((.((.((.(((.............))).)).....)).)))....
NM_020394_0      ---CACAGCAGAAUAUUUUUACUGAAUAAGAGUGUUACAAAUGUAAUGACUGUC---
NM_173531_40     CUUCACAGGAAAGCAUUUAUACUUGAGAAAUAUUGUACAAAUAUAAAGACUGUGAAA
NM_178558_16     -----------AGCAUUUUUACUUGAGAAAGAUUGUACAAGUAUAAAGAAUGUG---
```

4. Histone 3'UTR

5. NM020931: KIAA1586
   NM025079: zinc finger CCCH-type containing 12A (ZC3H12A), angiogenesis, apoptotic process, cell differentiation
   NM182508: FAM216B

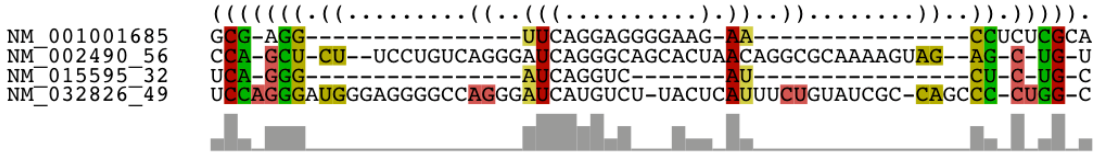
```
                 .(((((((((((((((((((((.((((((...)))).)))))).....))))))))))))))))).
NM_020931_16     U---------UGGCCCAUGUUU-G--------CU------AG----AGUGGGUCA---------U
NM_025079_24     GGAUUGGUUCUGGCCCAUGCAGCAC-CUCUAGCU---GUCUGCCUCAGUGGGUCAGAAGCGAUCA
NM_182508_0_     ACAGGAAGU-UUGCCCAUGUAU-CCCUGGUAUCUAGUGGGUGC---UUUGUGCA-UAUUUGUUGA
```

6. NM001077269: WAS/WASL interacting protein family (WIPF1), actin cytoskeleton
   NM001080432: fat mass obesity associated (FTO), oxidative DNA/RNA methylation, repair, ferrous iron binding
   NM002558: purinergic receptor P2X, ligand-gated ion channel (P2RX1), ion transport, signal transduction
   NM015888: hook homolog 1 (Drosophila) (HOOK1), microtubule cytoskeleton, actin binding
   NM152412: zinc finger protein 572 (ZNF572), regulation of transcription
   NM173829: SREK1-interacting protein (SREK1IP1), splicing



7. Cluster970: (same as cluster 8 above)

8. Cluster3946:
   NM001001685:FLJ45079
   NM002490: NADH dehydrogenase 1 alpha subcomplex (NDUFA6), mitochondrial electron transport, mitochondrial membrane
   NM015595: Rho guanine nucleotide exchange factor 26 (ARHGEF26), regulation of Rho protein signal transduction
   NM032826: solute carrier family 35 (SLC35B4), transmembrane transport



9. Cluster1501: (same as cluster 3 above)

10. Cluster1169
   NM001004349: FLJ45422
   NM001025072: chromosome 3 open reading frame 17
   NM007189: ATP-binding cassette subfamily F, member 2 (ABCF2)

Cut at the deepest depth to preserve the histone 3'UTR cluster

1. NM001008401: zinc finger protein 761 (ZNF761), regulation of transcription
   NM003429: ZNF85, transcription corepressor
   NM003441: ZNF141, regulation of transcription, anatomical structure morphogenesis
   NM006969: ZNF28, regulation of transcription
   NM015852: ZNF117, regulation of transcription
   NM021269: ZNF708, regulation of transcription
   NM025189: ZNF430, regulation of transcription
   NM133473: titin (TTN), adult heart development, ankyrin binding, ATP binding, protein tyrosine kinase activity