

Supplementary information:

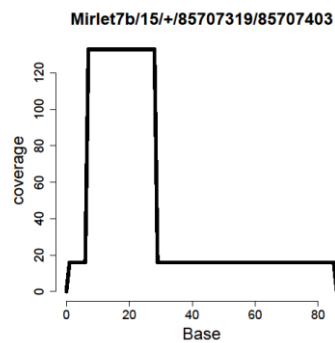
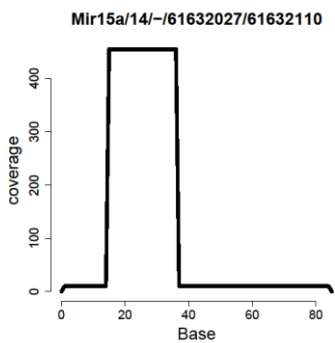
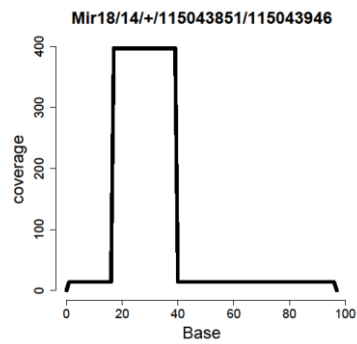
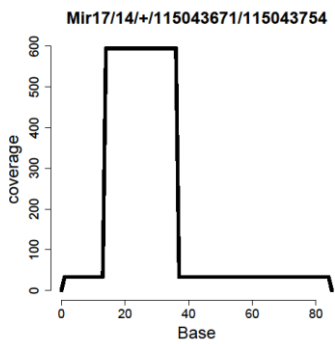
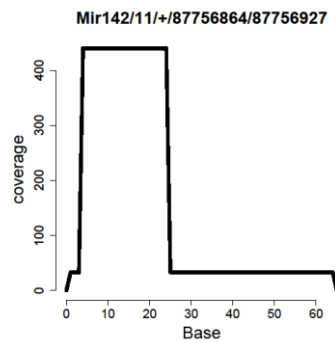
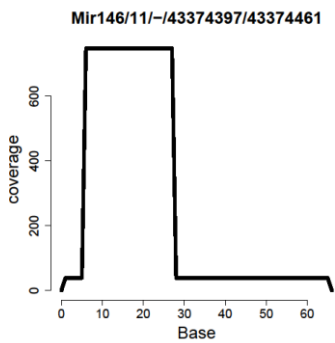
“SHARAKU: An algorithm for aligning and clustering read mapping profiles of deep sequencing in non-coding RNA processing”

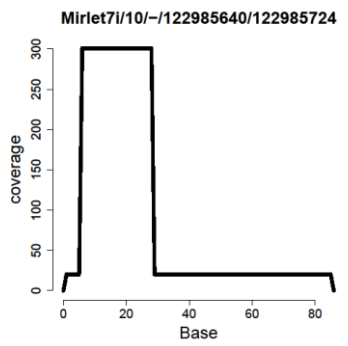
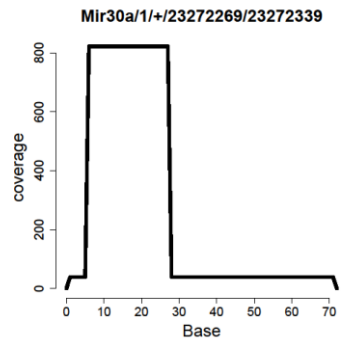
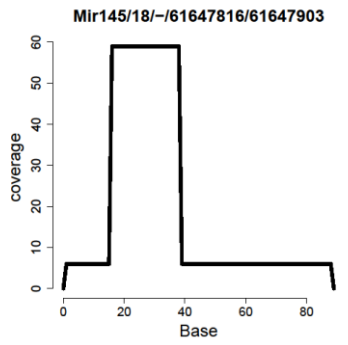
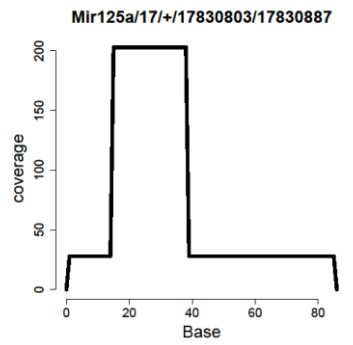
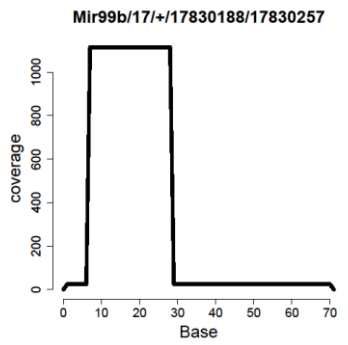
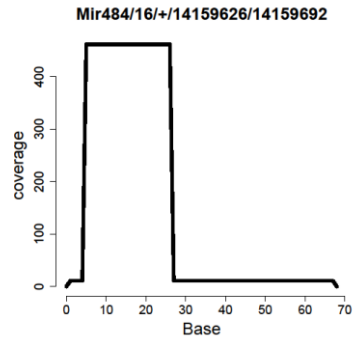
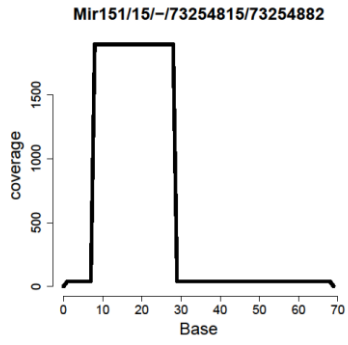
Mariko Tsuchiya, Kojiro Amano, Masaya Abe, Misato Seki, Sumitaka Hase, Kengo Sato and Yasubumi Sakakibara*

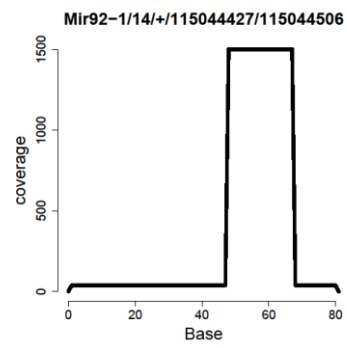
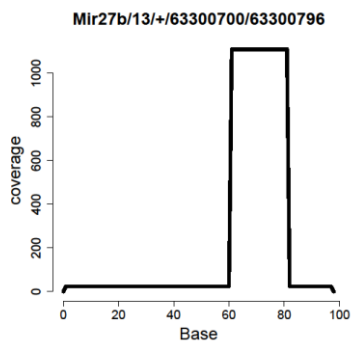
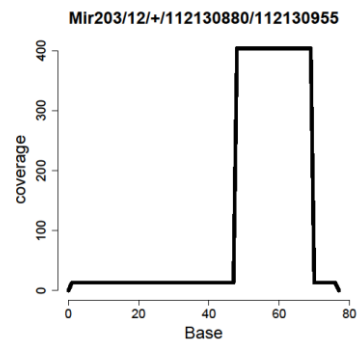
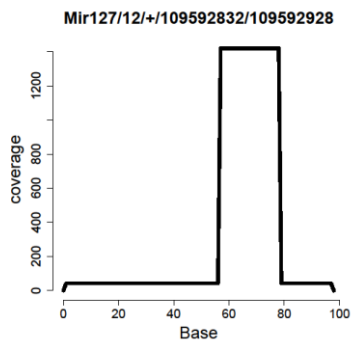
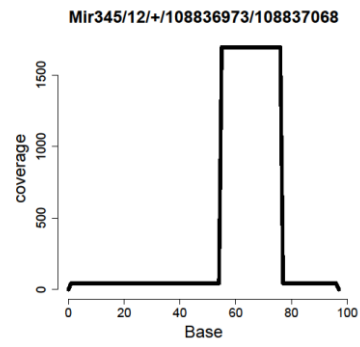
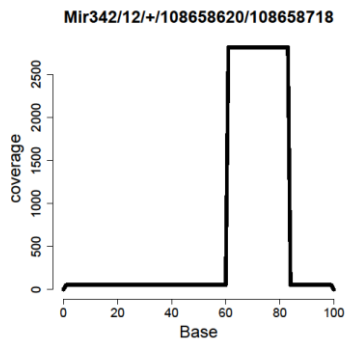
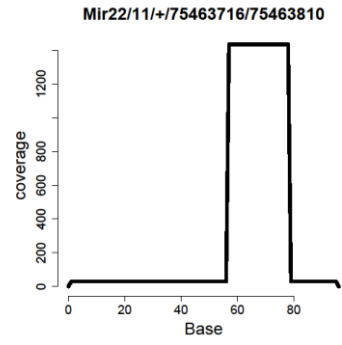
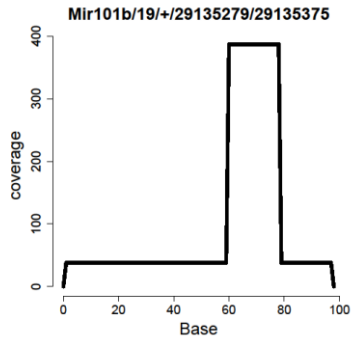
Supplemental Figure

Supplemental Figure 1. All read mapping profiles obtained by mapping the simulated read data to the 102 annotated small RNA sequences.

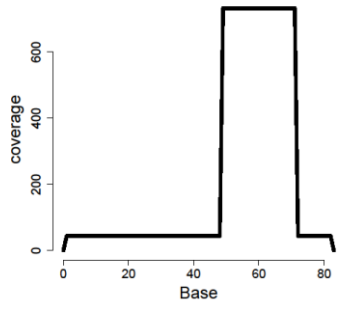
miRNAs (24 genes):



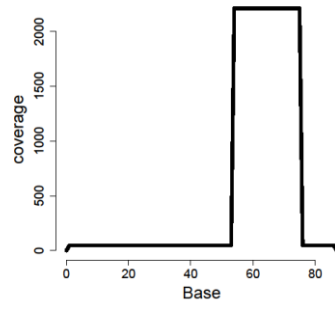




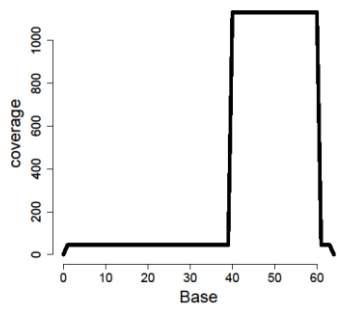
Mir19a/14/+/115044000/115044081



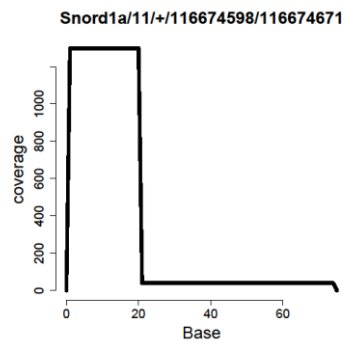
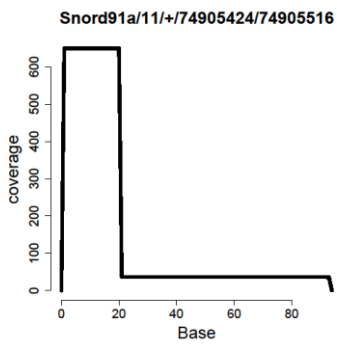
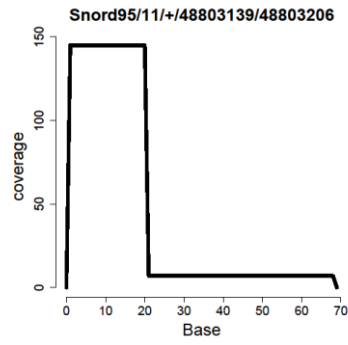
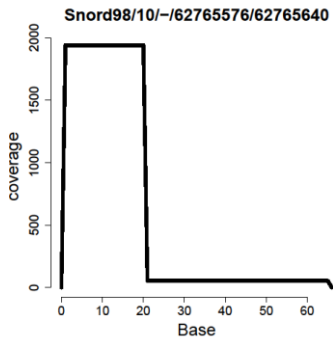
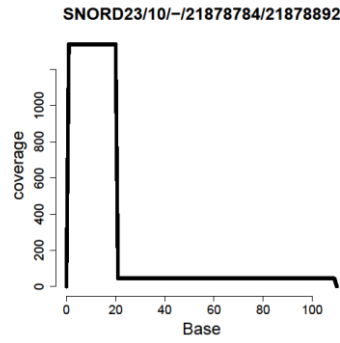
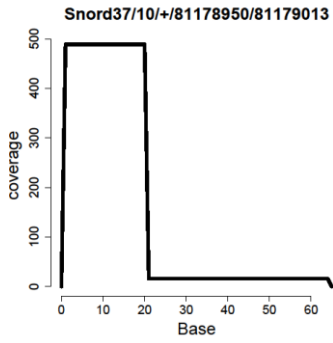
Mir28/16/+/24827855/24827940

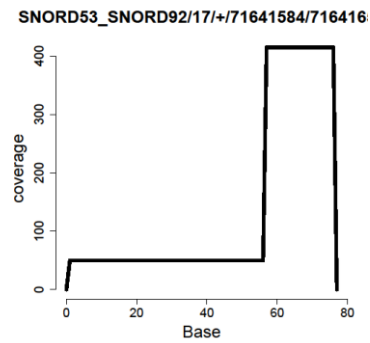
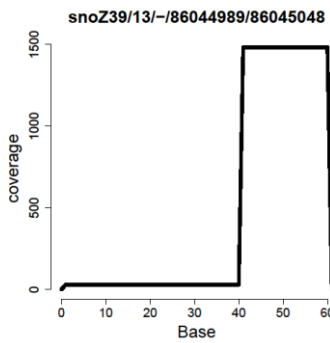
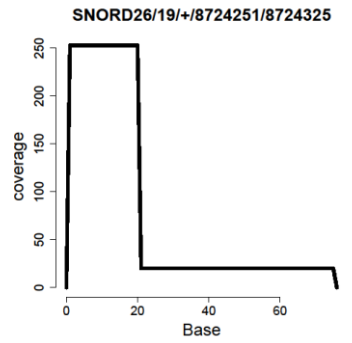
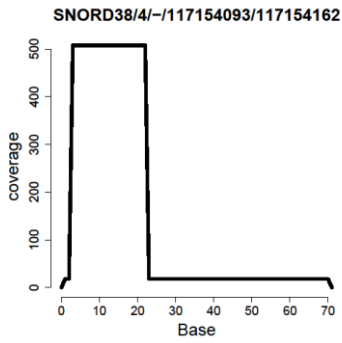
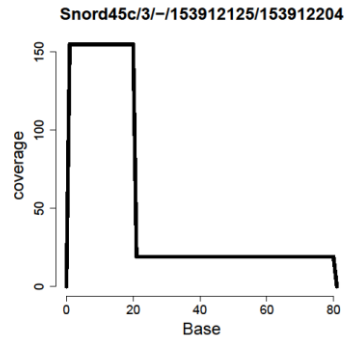
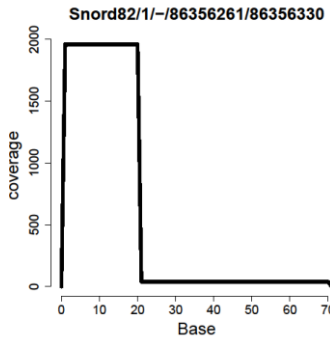
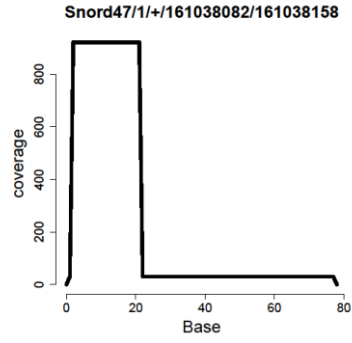
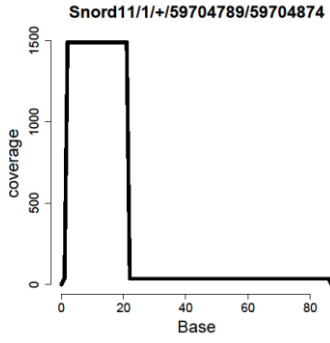


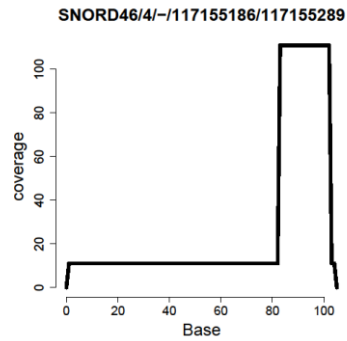
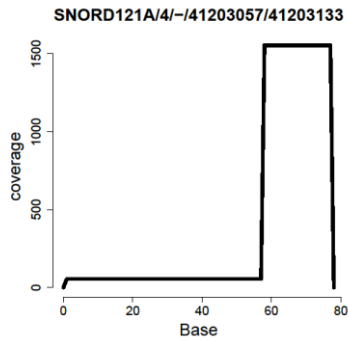
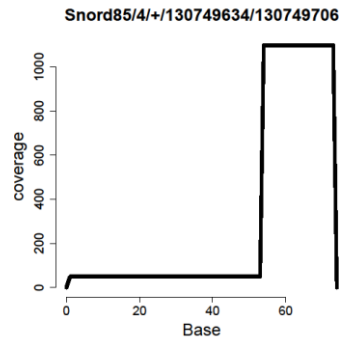
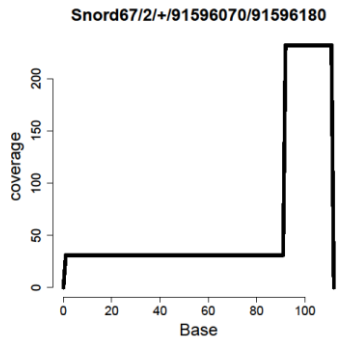
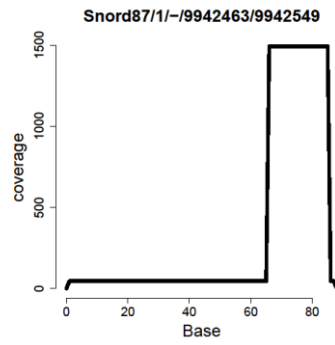
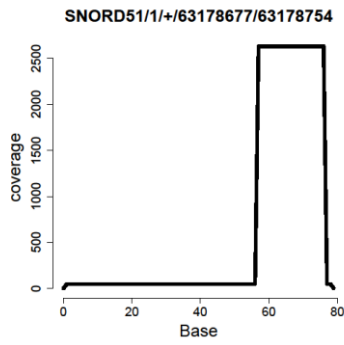
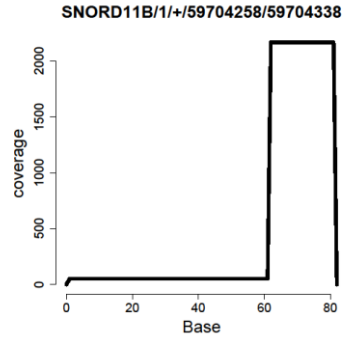
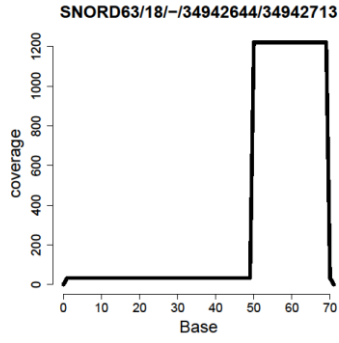
Mir143/18/-/61649196/61649258



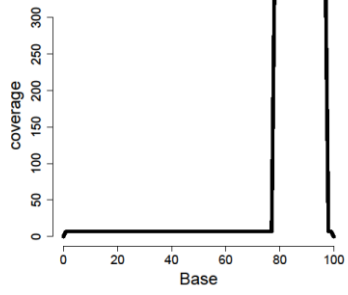
snoRNAs (37 genes):



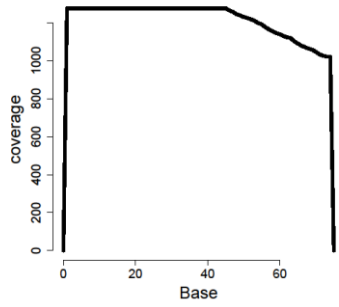




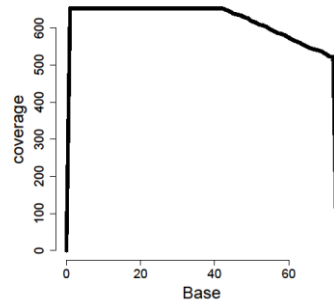
SNORD16/9/+/64175423/64175521



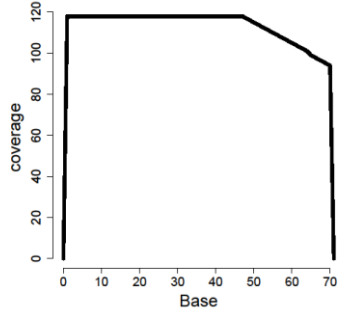
SNORD59/10/+/128084143/128084216



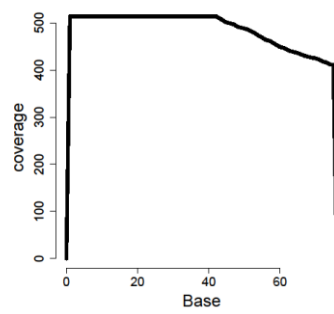
SNORD113/12/+/109656345/109656416



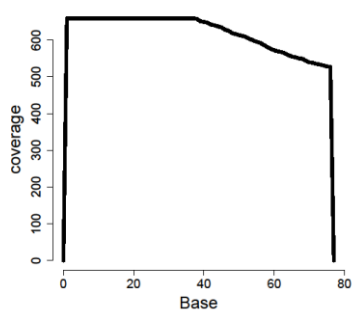
SNORD58/14/+/52069300/52069369



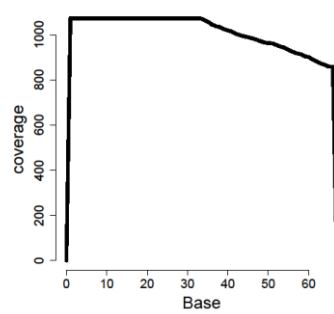
Snord66/16/+/20684239/20684313

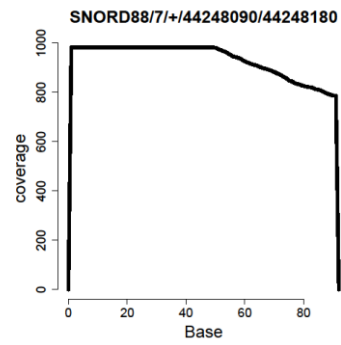
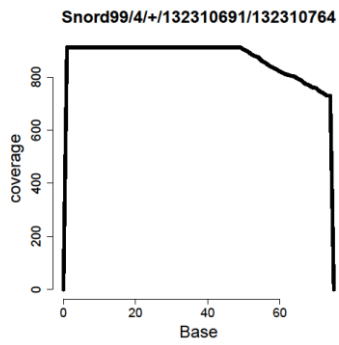
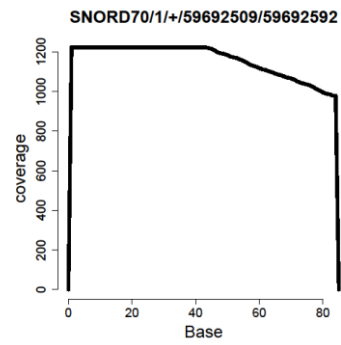
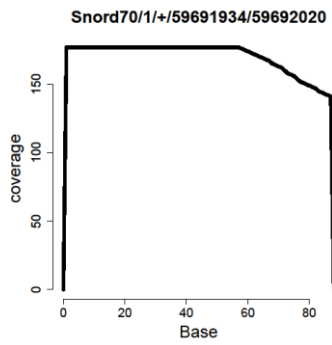
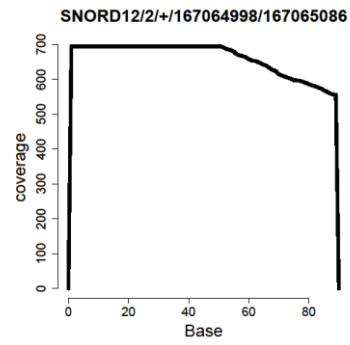
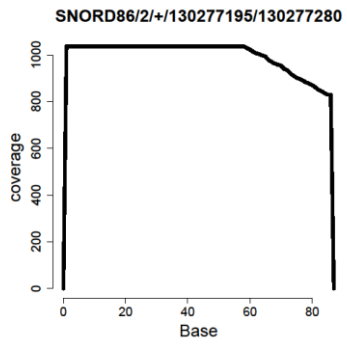
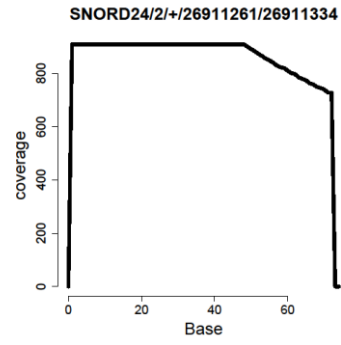
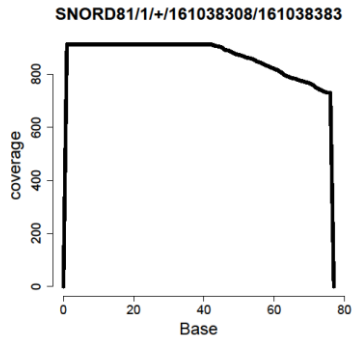


SNORD83/17/+/35247175/35247250

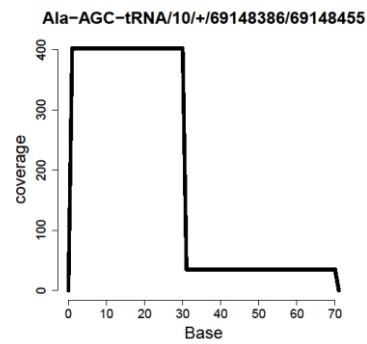
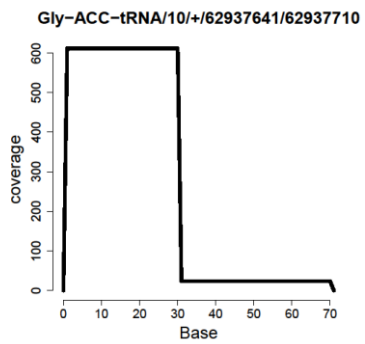
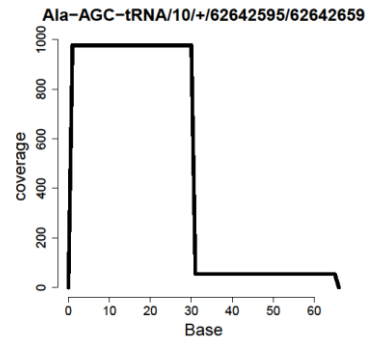
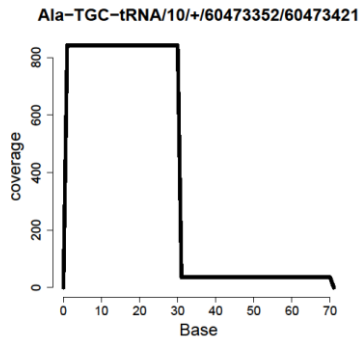
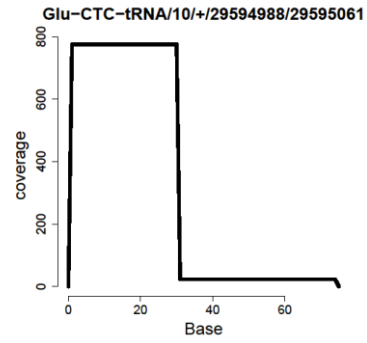
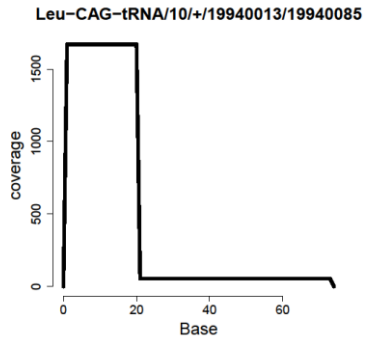


Snord58b/18/+/75001058/75001123

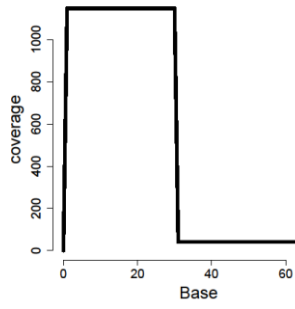




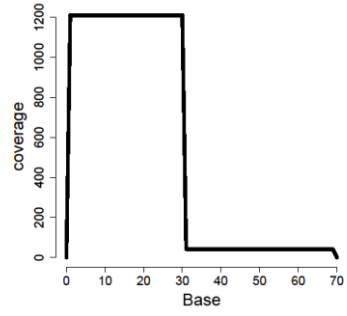
tRNAs (41 genes):



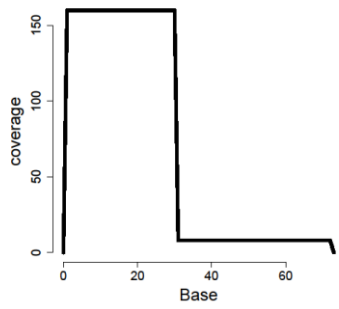
Ser-GCT-tRNA/10/+87075251/87075322



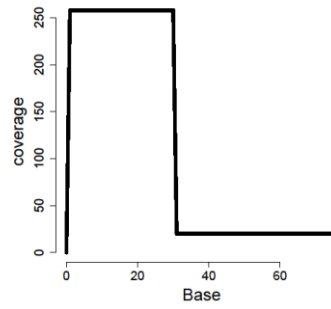
Ala-AGC-tRNA/10/+95063309/95063377



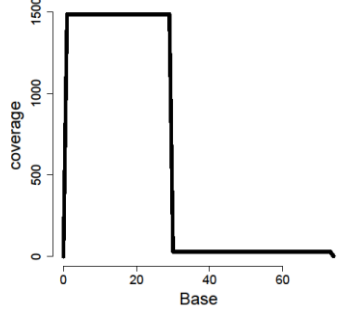
Ser-GCT-tRNA/10/+103356103/103356174



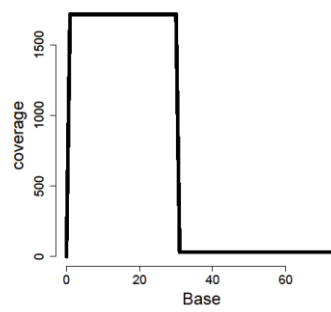
Ile-AAT-tRNA/10/+105391053/105391127



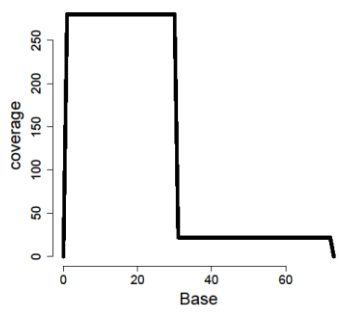
Asn-GTT-tRNA/10/+110810401/110810472



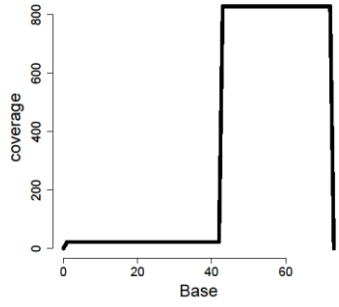
Ala-AGC-tRNA/10/+111109450/111109521



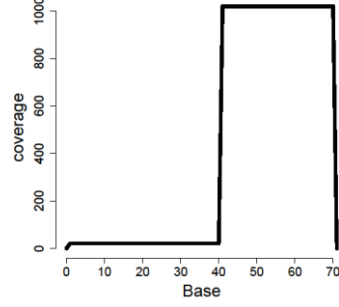
Thr-GGT-tRNA/10/+126900102/126900173



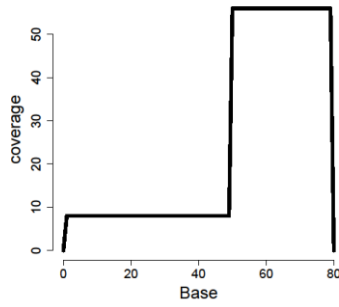
Met-CAT-tRNA/12/+86551561/86551632



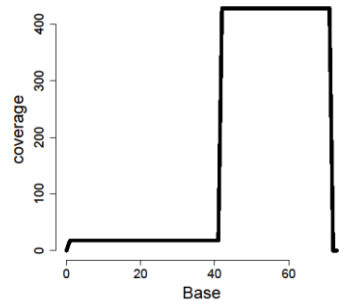
Ser-GCT-tRNA/11/-95952589/95952658



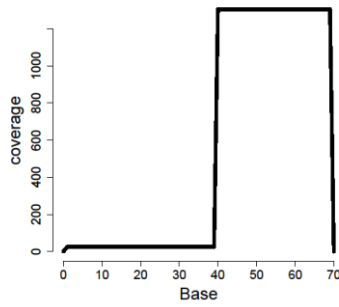
Gly-ACC-tRNA/11/-105247999/105248071



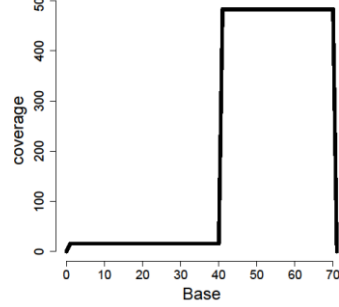
Ala-AGC-tRNA/11/-116415979/116416051



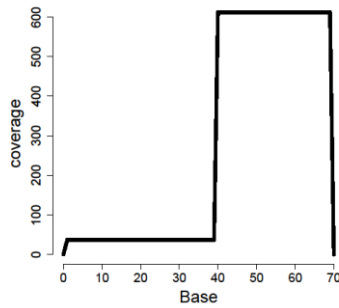
Thr-TGT-tRNA/11/-117987594/117987666



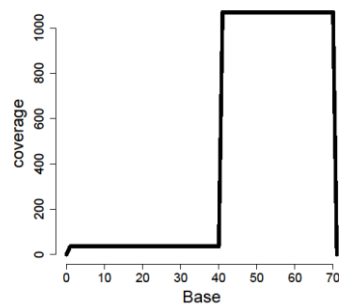
Ala-TGC-tRNA/12/+16321418/16321487



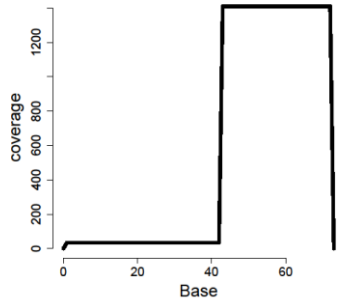
Ala-AGC-tRNA/12/+16852070/16852139



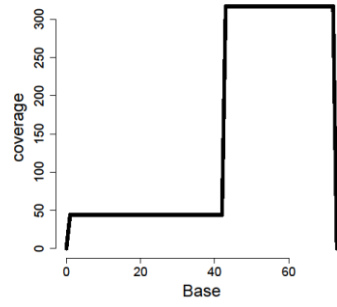
Glu-CTC-tRNA/12/+24467219/24467290



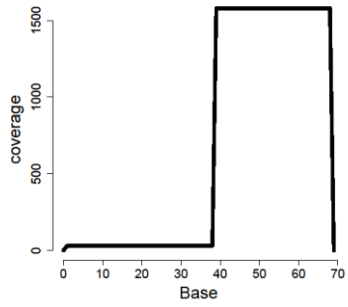
Val-AAC-tRNA/12/+31651885/31651956



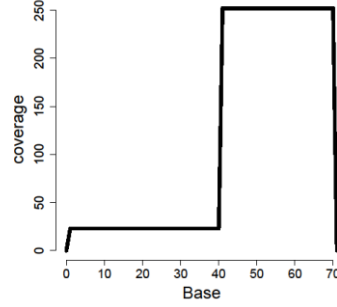
Ala-AGC-tRNA/12/+54244071/54244142



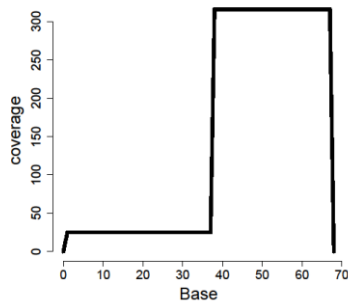
Asn-GTT-tRNA/12/+66833318/66833385



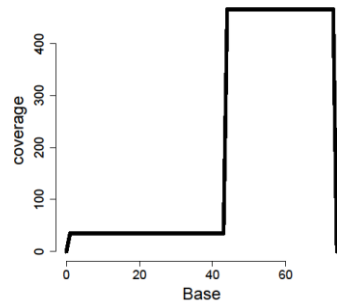
Glu-CTC-tRNA/12/+73003789/73003861



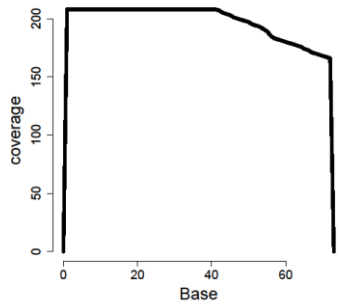
Ala-AGC-tRNA/12/+91730713/91730781



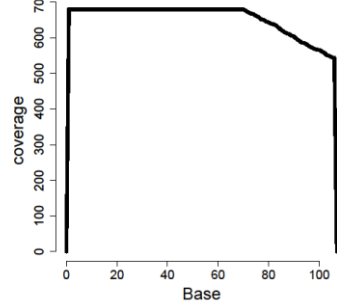
Arg-CCT-tRNA/6/+38533931/38534003

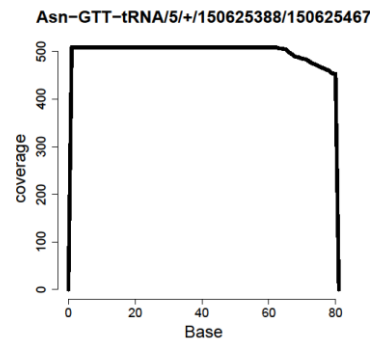
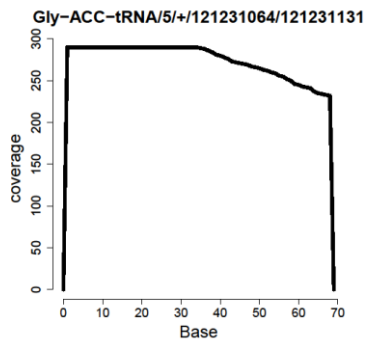
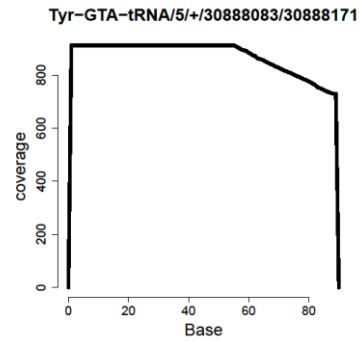
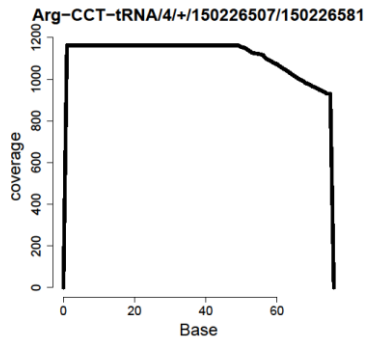
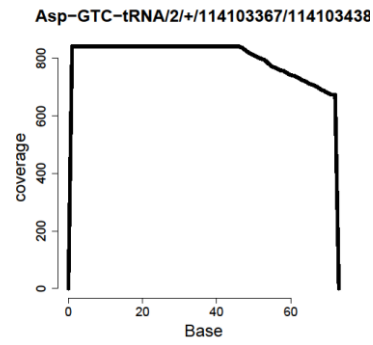
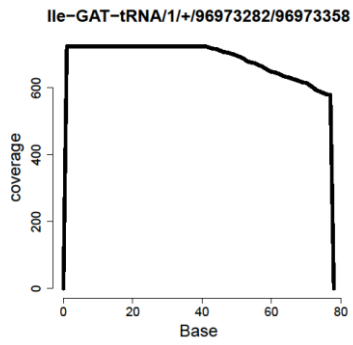
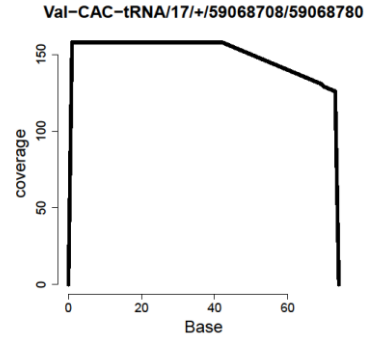
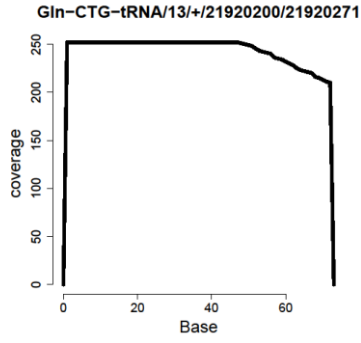


Phe-AAA-tRNA/11/+94888957/94889028

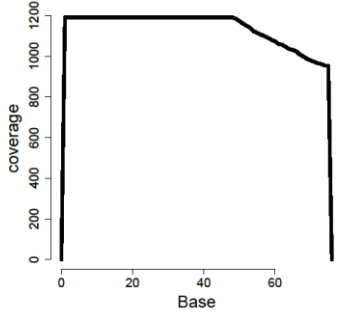


Leu-CAA-tRNA/13/+21914838/21914943

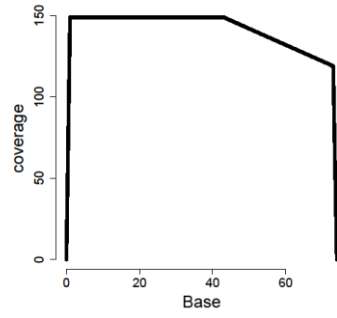




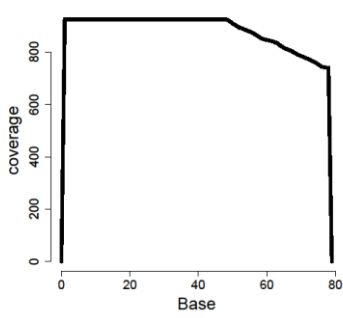
Lys-CTT-tRNA/6/+/125308431/125308505



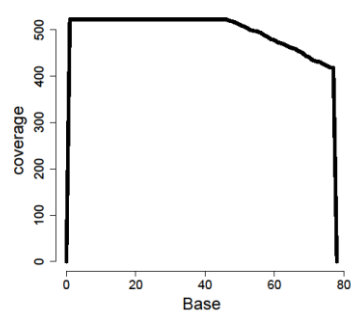
Gly-TCC-tRNA/8/+/36717443/36717515



Asn-GTT-tRNA/8/+/120625978/120626055

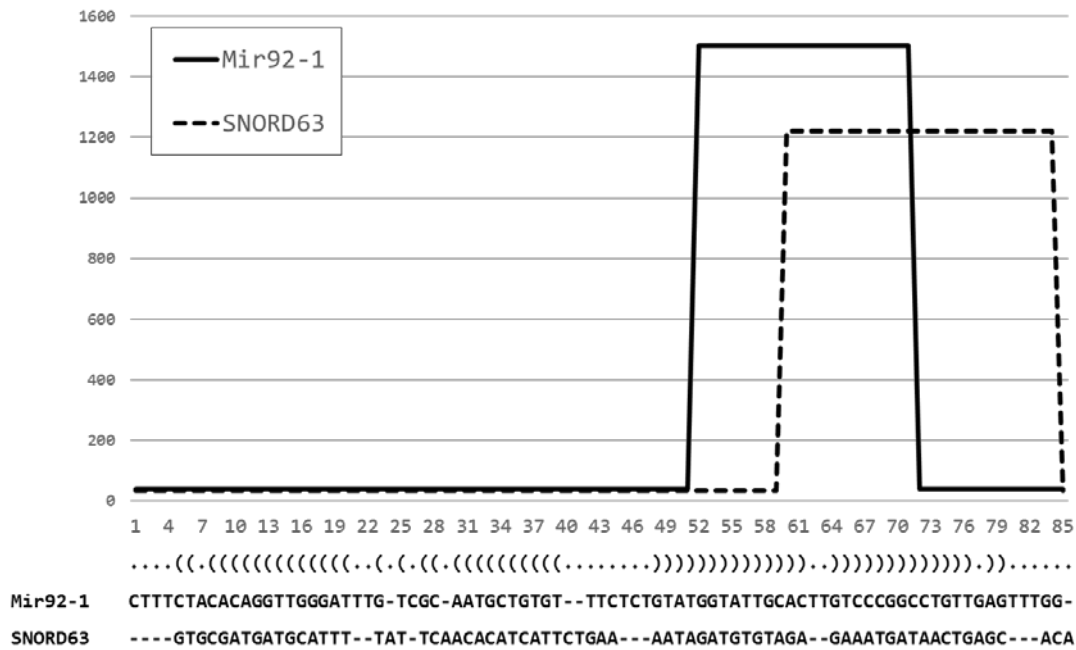


Val-AAC-tRNA/9/+/77393736/77393812



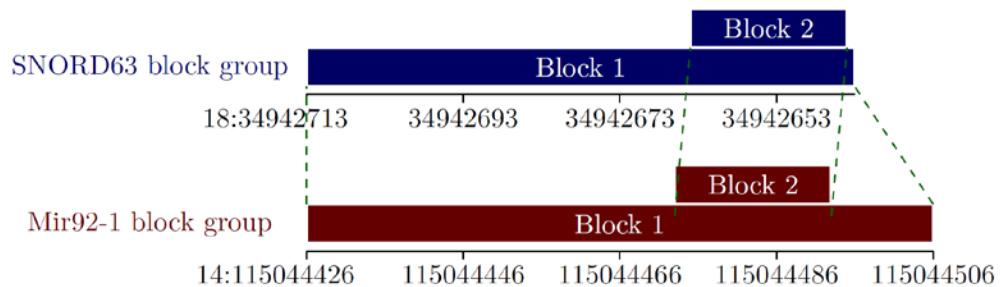
Supplemental Figure 2. (a) An alignment produced by SHARAKU for a pair of RNAs (Mir92-1 and SNORD63) in the simulated dataset with the annotations of RNA sequence alignment and the predicted secondary structure. The solid line represents the read coverages of Mir92-1, and the dashed line represents the read coverages of SNORD63. The similarity score is 0.562. (b) An alignment produced by deepBlockAlign for a pair of RNAs (Mir92-1 and SNORD63) in the simulated dataset. The similarity score is 0.87.

(a)



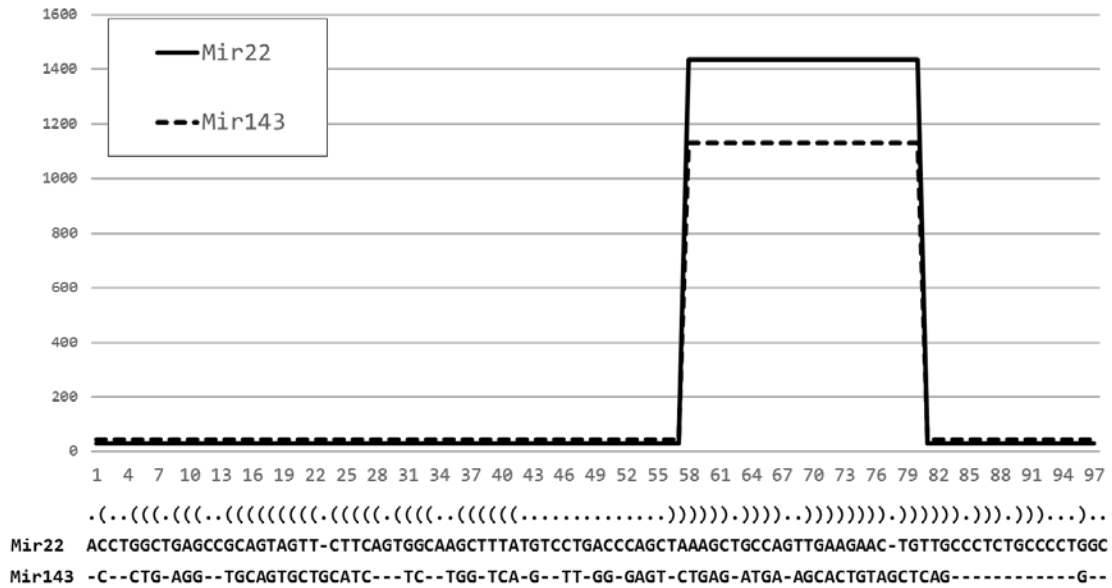
(b)

cluster-64-Mir92-1(+) with cluster-77-SNORD63(-) [Score: 0.87]



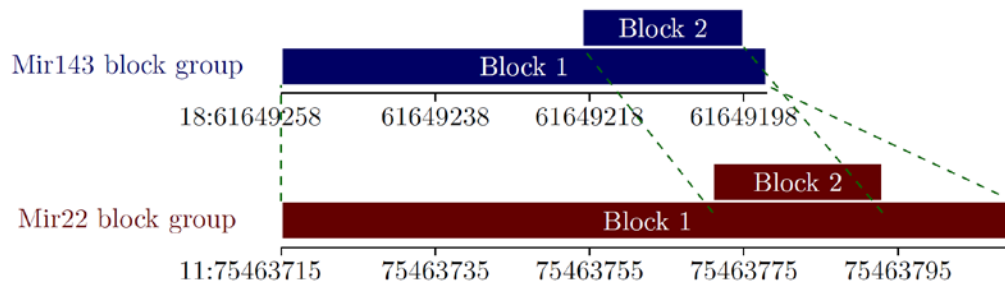
Supplemental Figure 3. (a) An alignment produced by SHARAKU for a pair of RNAs (Mir22 and Mir143) in the simulated dataset with the annotations of RNA sequence alignment and the predicted secondary structure. The solid line represents the read coverages of Mir22, and the dashed line represents the read coverages of Mir143. The similarity score is 0.999. (b) An alignment produced by deepBlockAlign for a pair of RNAs (Mir22 and Mir143) in the simulated dataset. The similarity score is 0.34.

(a)



(b)

cluster-33-Mir22(+) with cluster-79-Mir143(-) [Score: 0.34]

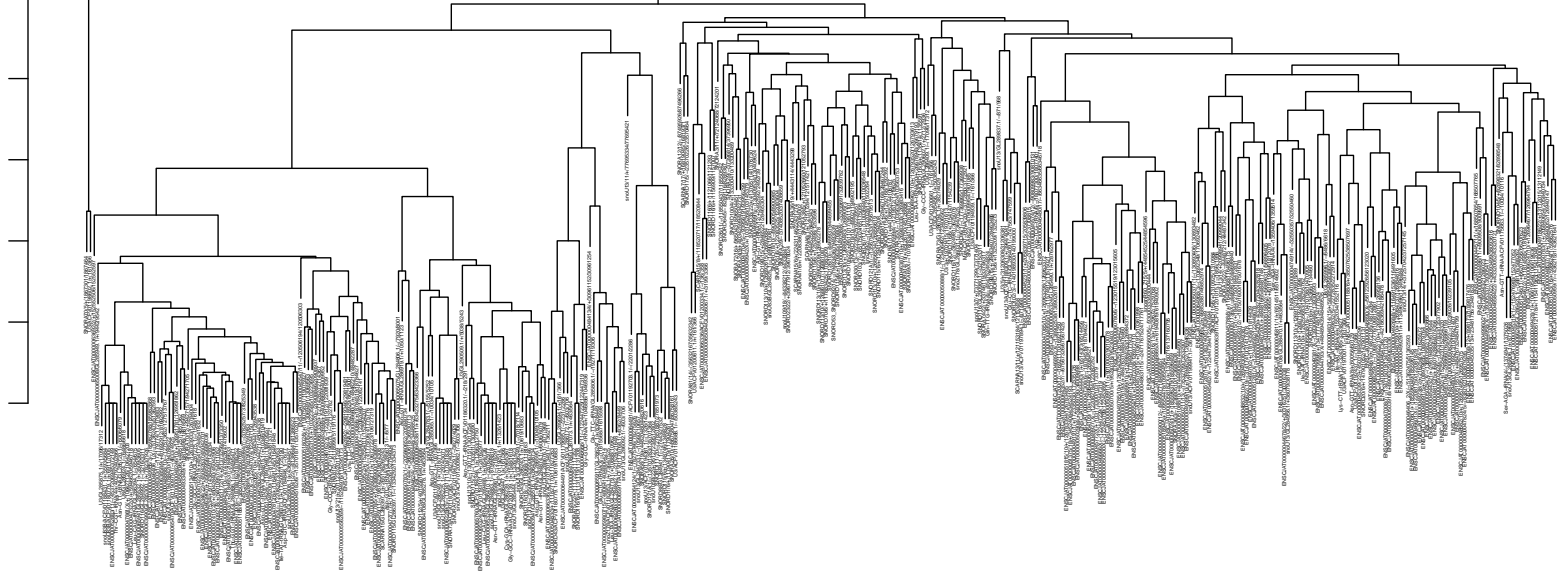


Supplemental Figure 4.

(a) Dendrogram of the hierarchical clustering result based on deepBlockAlign alignments for the read mapping profiles on the real dataset obtained from an RNA-seq experiment of the hippocampus of common marmoset.

As the input to deepBlockAlign, the output of blockbuster executed on the input of the real dataset was given together with the sequence annotation information. The blocks with the maximum read coverage less than 3 were all removed.

0.0 0.2 0.4 0.6 0.8 1.0



(b) Dendrogram of the hierarchical clustering result based on SHARAKU alignments for the read mapping profiles of 619 non-coding RNAs.

The annotated RNA sequences with the maximum read coverage less than 20 were all removed.

