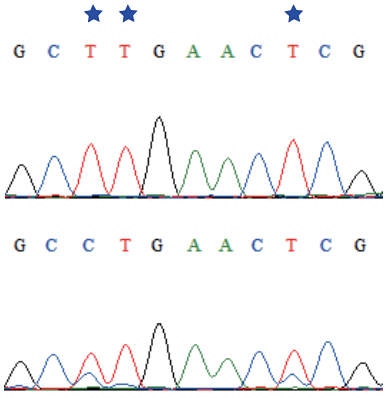
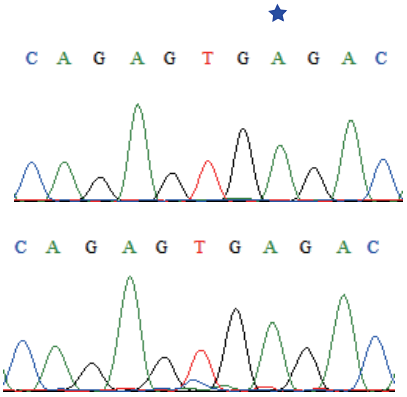
Supplementary Materials

Supplementary fig. S1. Results of Sanger sequencing validation for 78 randomly-selected RNA editing sites.

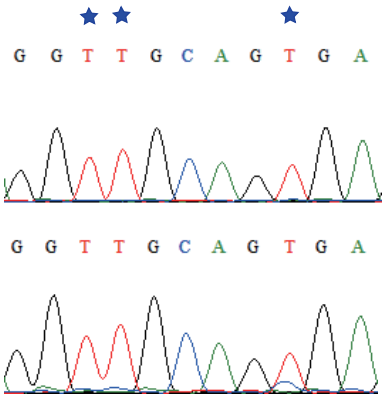
chr1:9227372:A-to-G:+
chr1:9227373:A-to-G:+
chr1:9227378:A-to-G:+



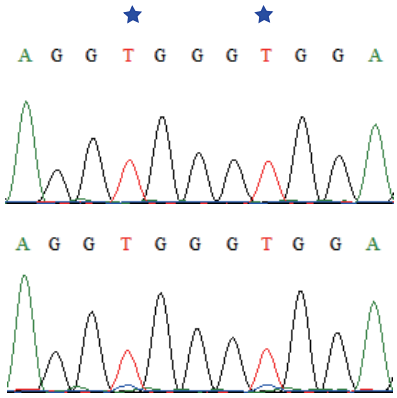
chr3:13204246:A-to-G:-



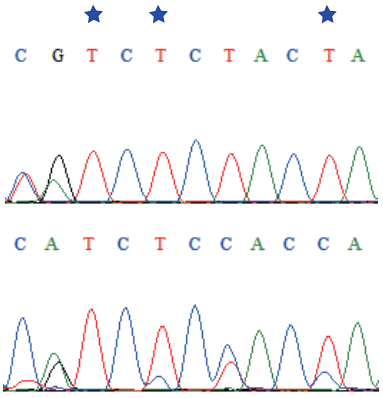
chr3:13204197:A-to-G:-
chr3:13204198:A-to-G:-
chr3:13204203:A-to-G:-



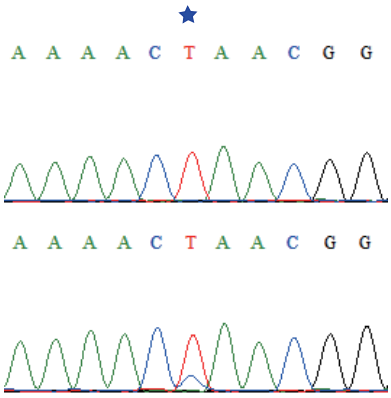
chr3:13204034:A-to-G:-
chr3:13204038:A-to-G:-



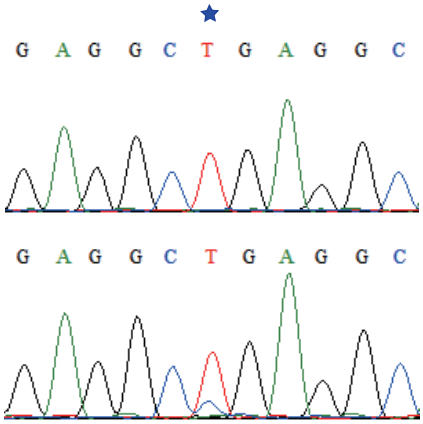
chr3:13204094:A-to-G:-
chr3:13204096:A-to-G:-
chr3:13204099:A-to-G:-



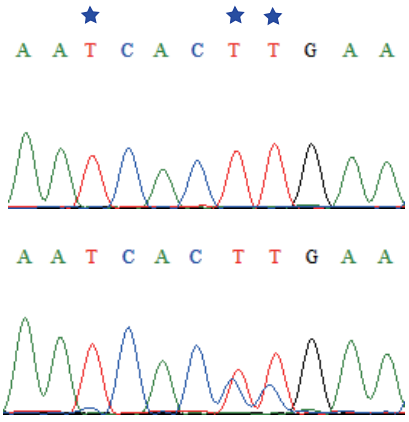
chr3:13204113:A-to-G:-



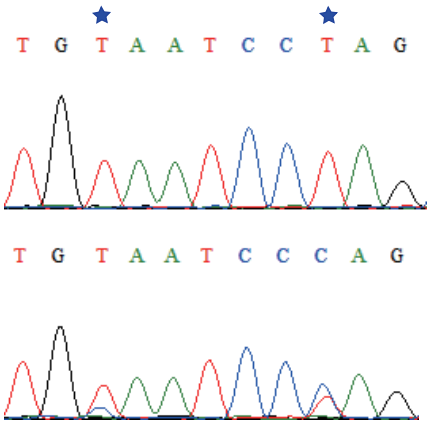
chr3:13204161:A-to-G:-



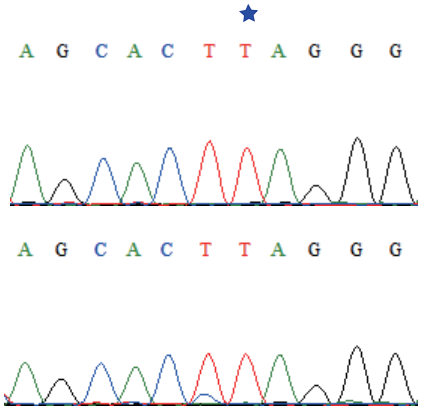
chr3:13204173:A-to-G:-
chr3:13204177:A-to-G:-
chr3:13204178:A-to-G:-



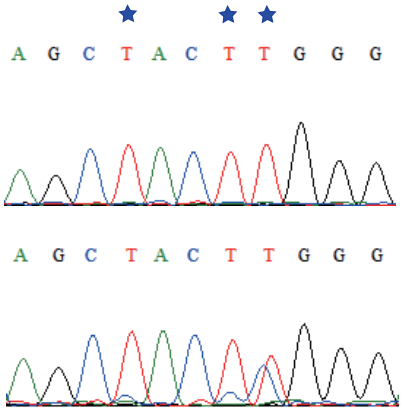
chr6:128380981:A-to-G:-
chr6:128380987:A-to-G:-



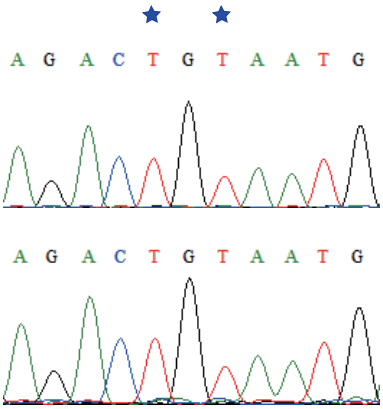
chr6:128380993:A-to-G:-



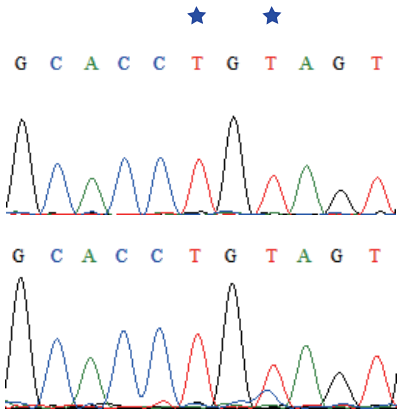
chr6:128381114:A-to-G:-
chr6:128381117:A-to-G:-
chr6:128381118:A-to-G:-



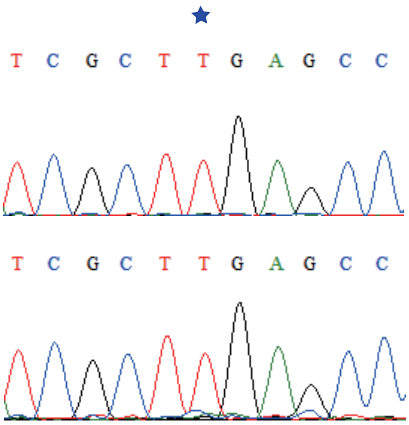
chr6:128381163:A-to-G:-
chr6:128381165:A-to-G:-



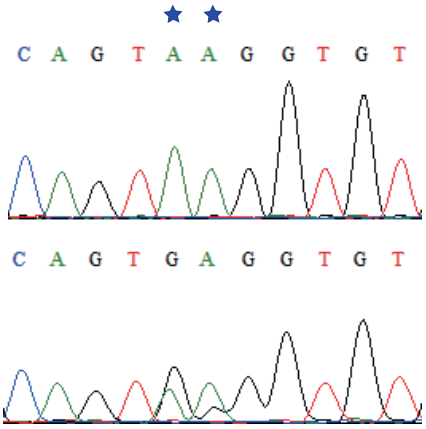
chr6:128381102:A-to-G:-
chr6:128381104:A-to-G:-



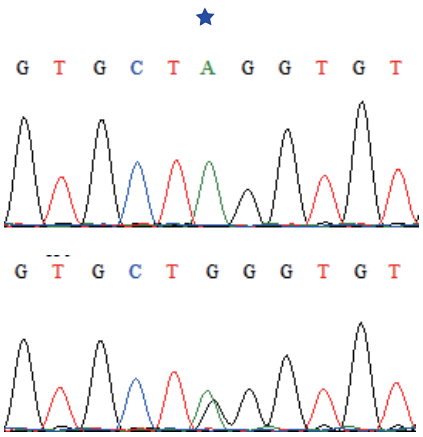
chr6:128381143:A-to-G:-



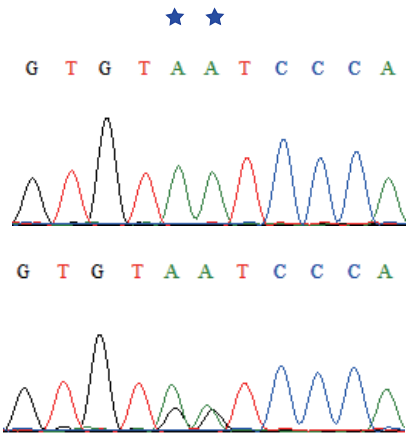
chr1:12321981:A-to-G:+
chr1:12321982:A-to-G:+



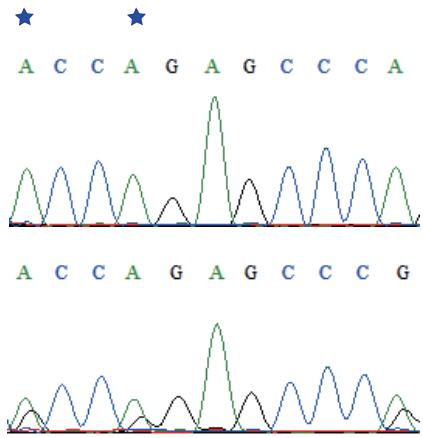
chr1:12321894:A-to-G:+



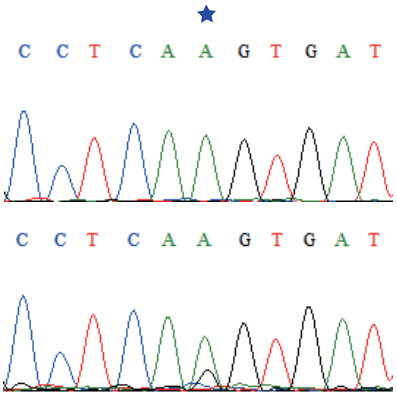
chr1:12321916:A-to-G:+
chr1:12321917:A-to-G:+



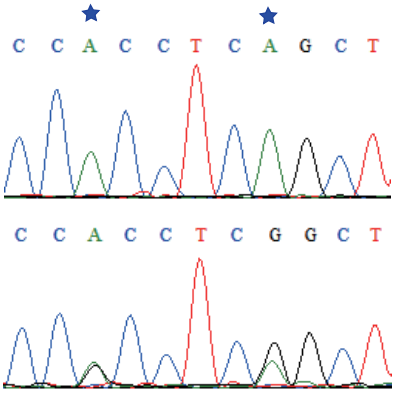
chr1:12321952:A-to-G:+
chr1:12321955:A-to-G:+
chr1:12321962:A-to-G:+



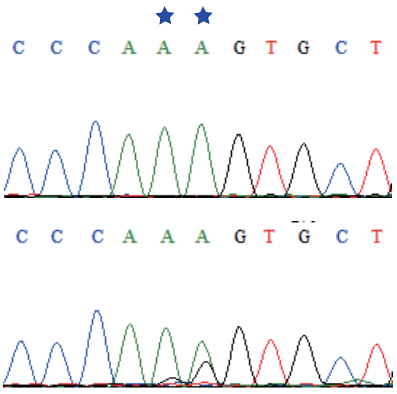
chr1:55676567:A-to-G:+



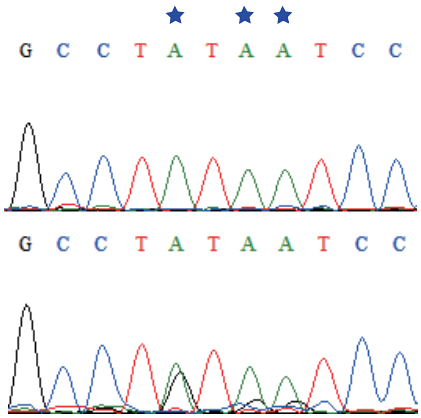
chr1:55676579:A-to-G:+
chr1:55676584:A-to-G:+



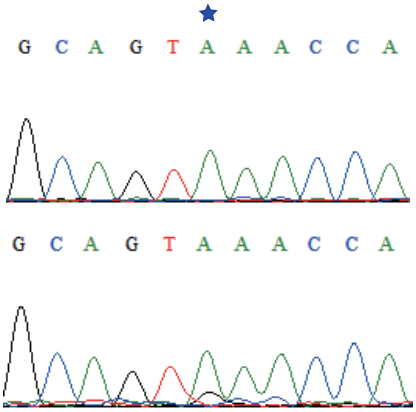
chr1:55676593:A-to-G:+
chr1:55676594:A-to-G:+



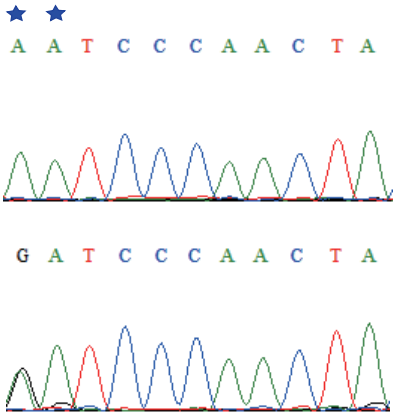
chr8:39450161:A-to-G:+
chr8:39450163:A-to-G:+
chr8:39450164:A-to-G:+



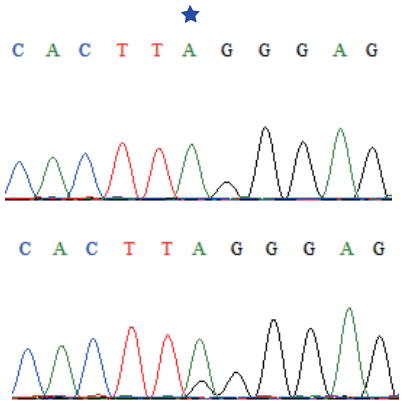
chr8:39450228:A-to-G:+



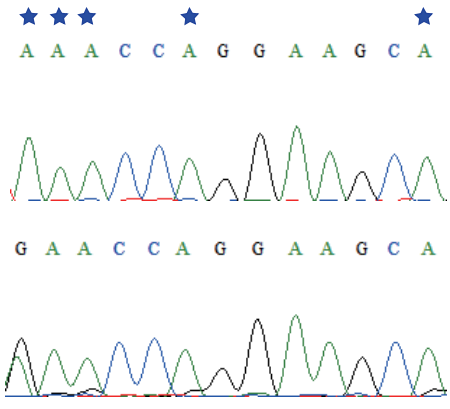
chr13:9563033:A-to-G:+
chr13:9563034:A-to-G:+
chr13:9563043:A-to-G:+



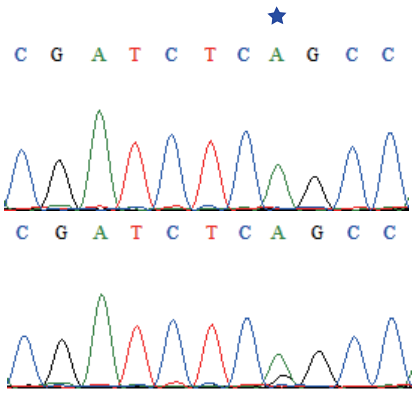
chr13:9562902:A-to-G:+



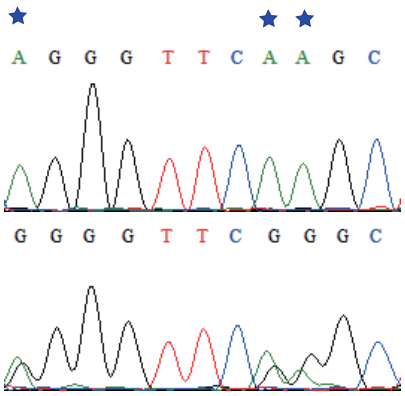
chr13:9563073:A-to-G:+
chr13:9563074:A-to-G:+
chr13:9563075:A-to-G:+
chr13:9563078:A-to-G:+
chr13:9563085:A-to-G:+



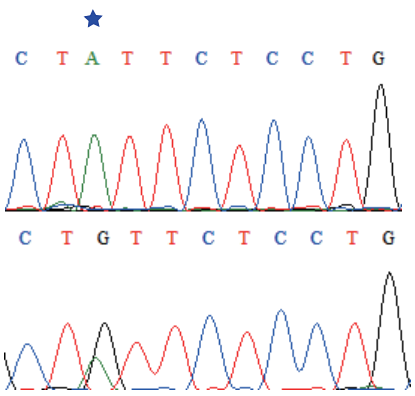
chr2:145905114:A-to-G:+



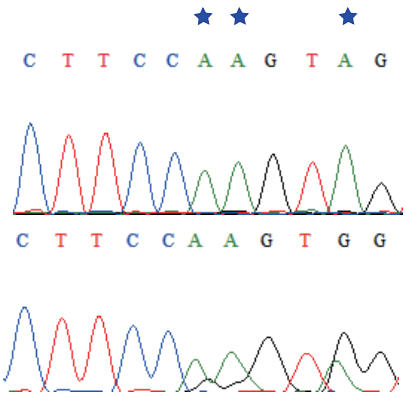
chr2:145905136:A-to-G:+
chr2:145905143:A-to-G:+
chr2:145905144:A-to-G:+



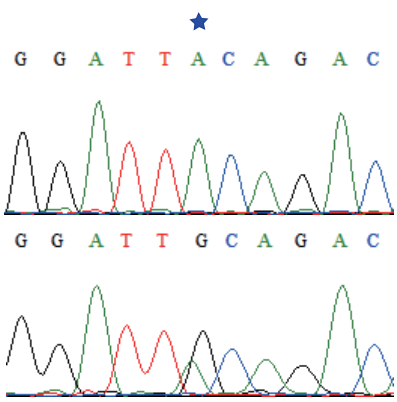
chr2:145905148:A-to-G:+



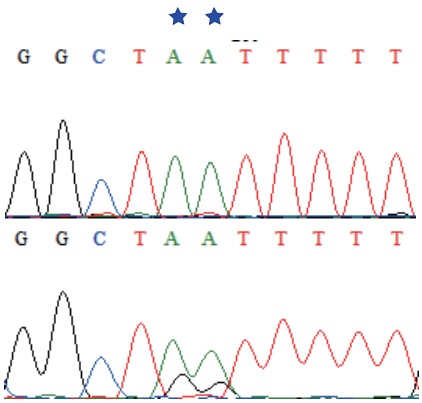
chr2:145905169:A-to-G:+
chr2:145905170:A-to-G:+
chr2:145905173:A-to-G:+



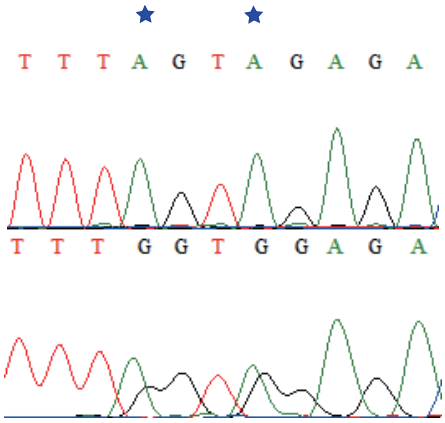
chr2:145905183:A-to-G:+



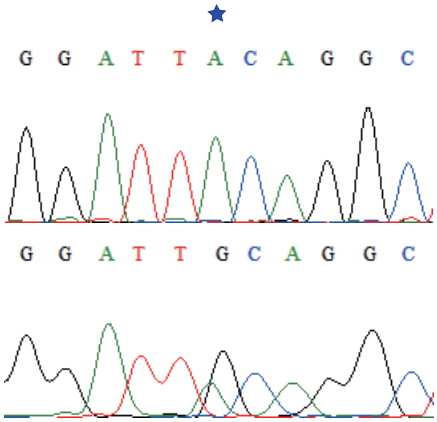
chr2:145905209:A-to-G:+
chr2:145905210:A-to-G:+



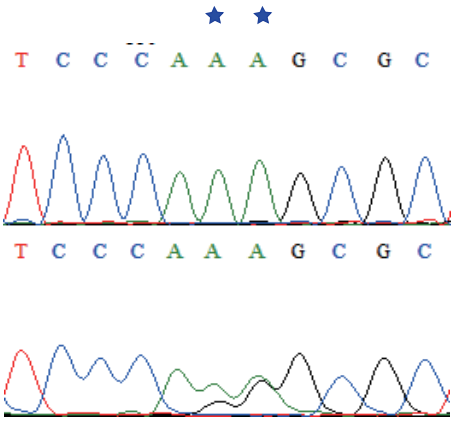
chr2:145905224:A-to-G:+
chr2:145905227:A-to-G:+



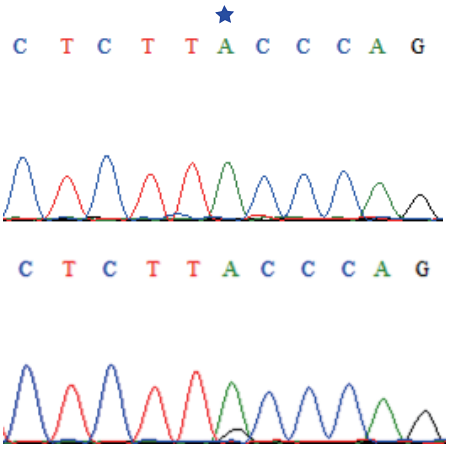
chr2:145905316:A-to-G:+



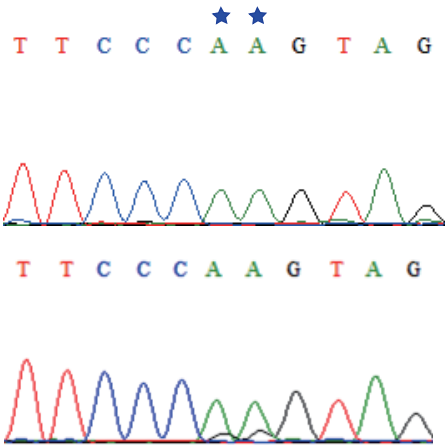
chr2:145905303:A-to-G:+
chr2:145905304:A-to-G:+



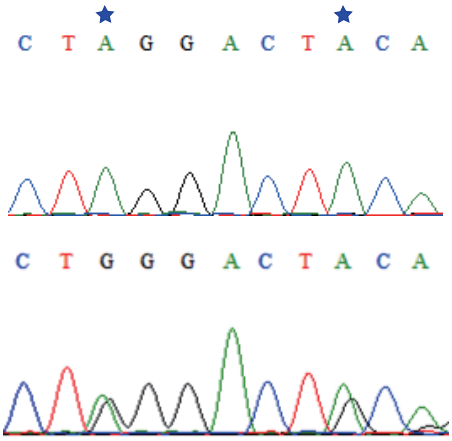
chr8:39451384:A-to-G:+



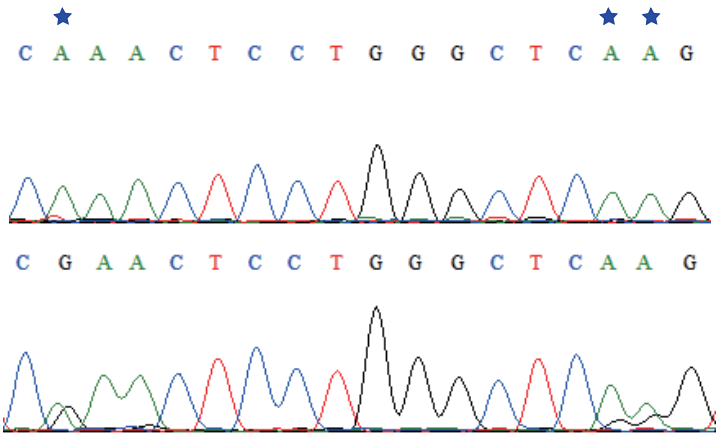
chr8:39451470:A-to-G:+
chr8:39451471:A-to-G:+

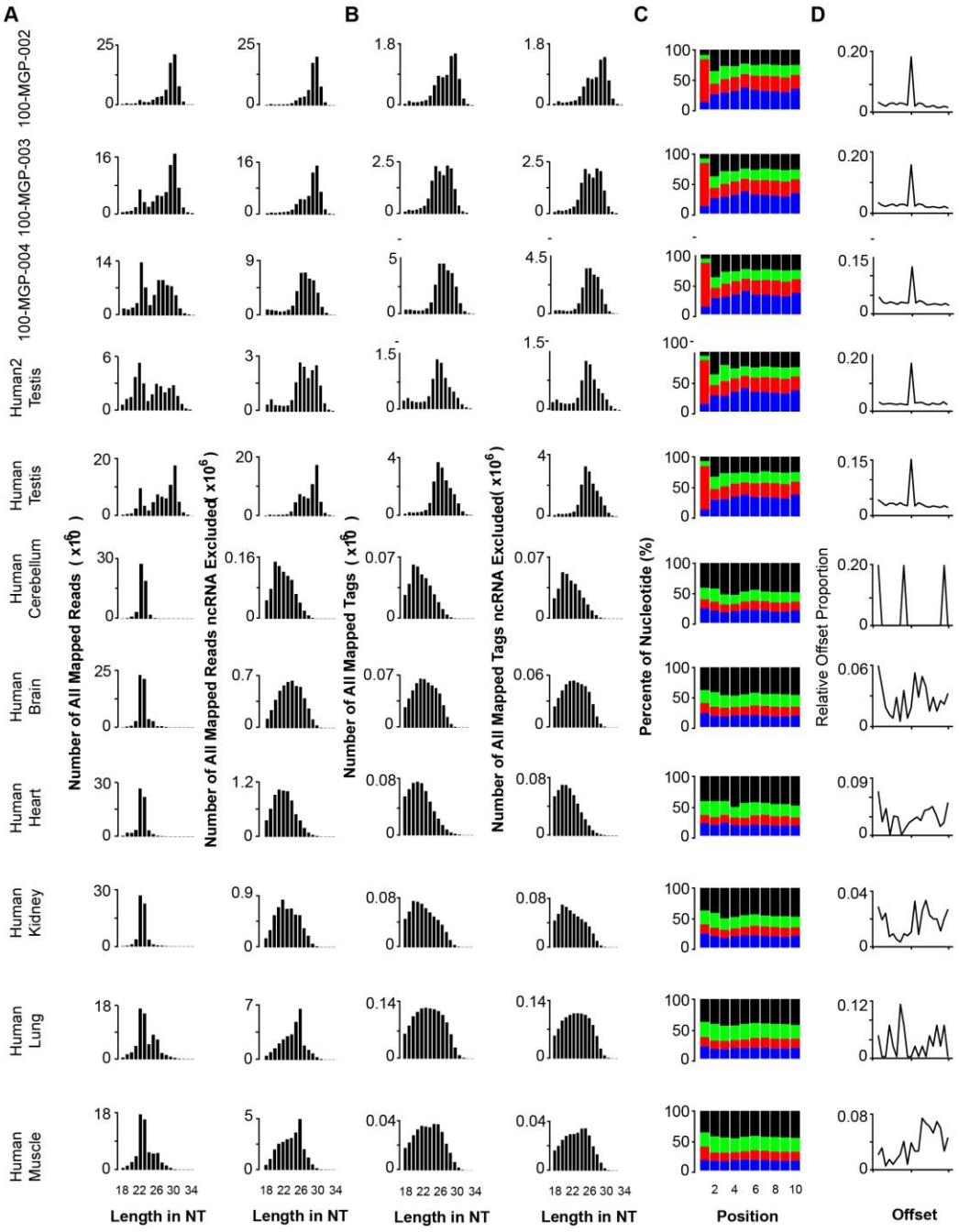


chr8:39451478:A-to-G:+
chr8:39451484:A-to-G:+
chr8:39451486:A-to-G:+

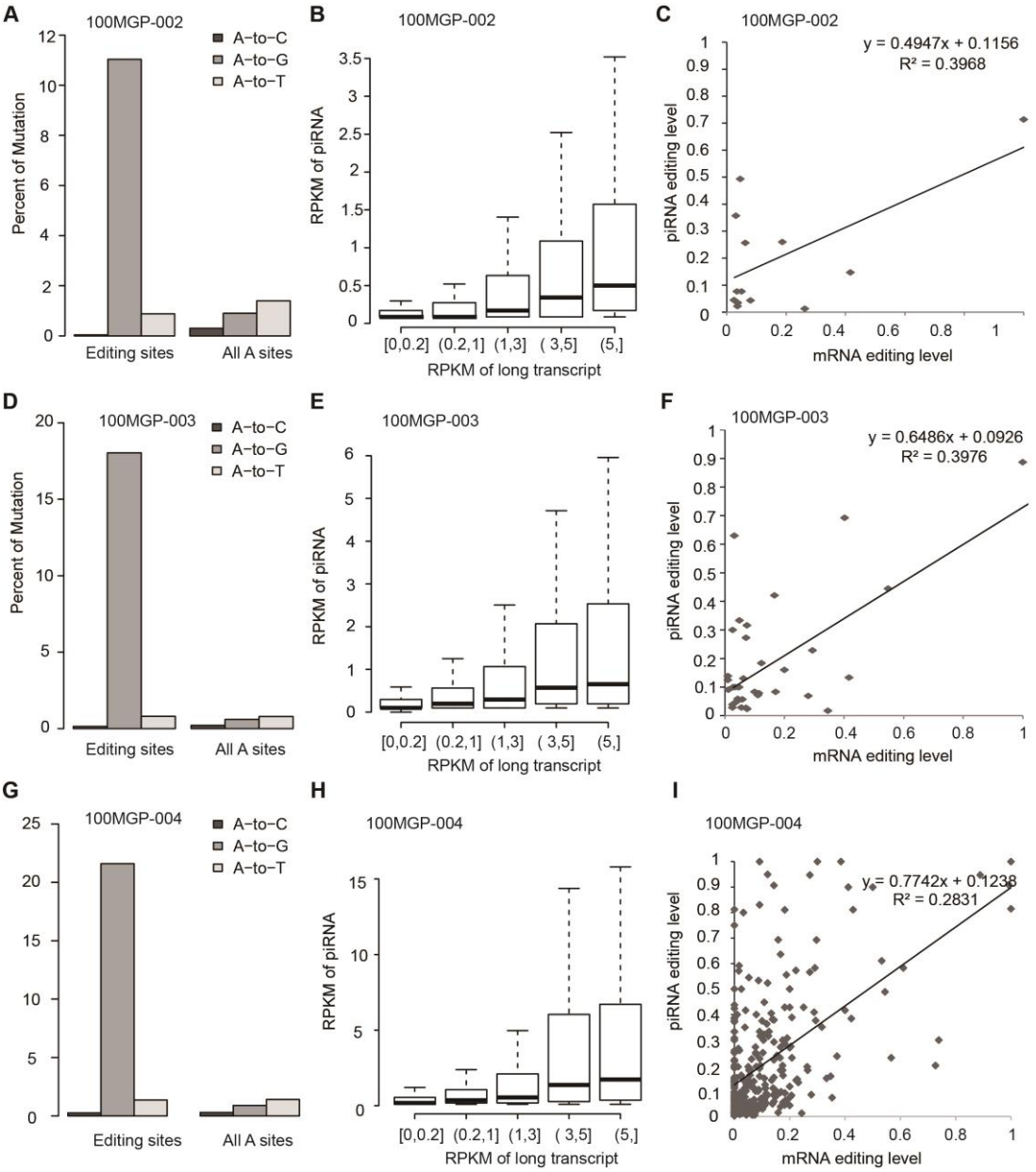


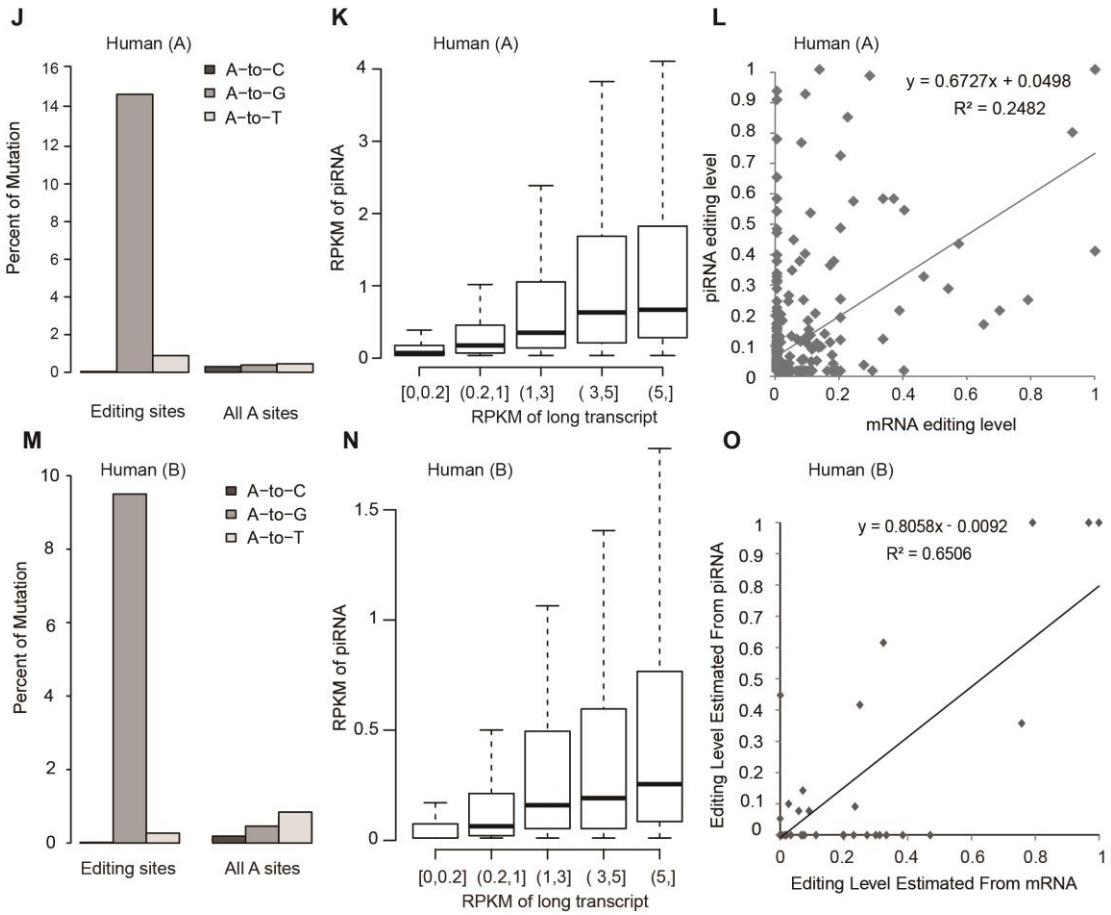
chr8:39451565:A-to-G:+
chr8:39451479:A-to-G:+
chr8:39451480:A-to-G:+



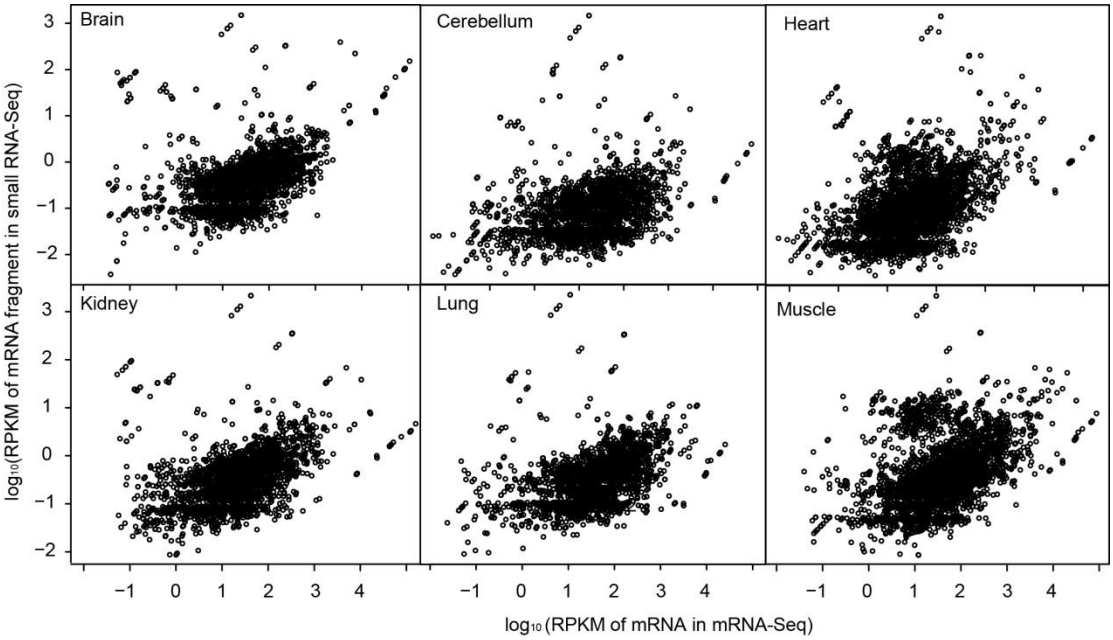


Supplementary fig. S2. Characteristics of piRNAs identified in three additional rhesus macaque animals and two human individuals (related to fig. 3). (**A, B**) Histograms showing the length distributions of mapped reads/tags before and after excluding annotated non-coding RNA in different tissues. **(C)** Nucleotide preference of the first 10nt from 5'-end of candidate piRNAs, A, U, C and G were marked in blue, red, green and black, respectively. **(D)** For each head-to-head overlapping piRNA pair, the complementary length was calculated, and the proportions of piRNAs with different complementary lengths are shown.

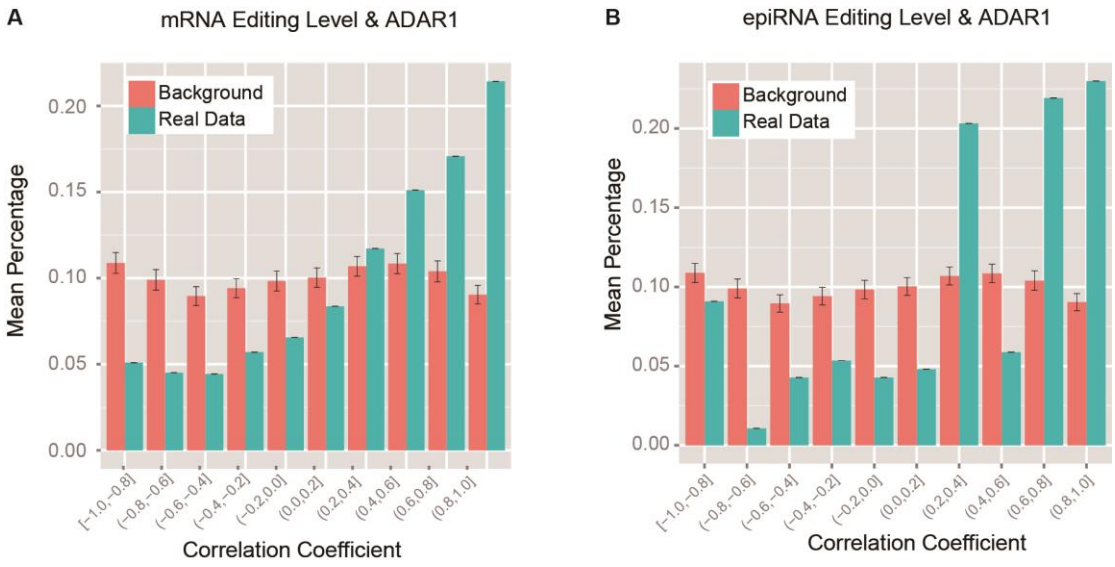

A
100MGP-002
A−to−C
A−to−G
A−to−T
Percent of Mutation
Editing sites
All A sites
B
100MGP-002
RPKM of piRNA
[0,0.2]
(0.2,1]
(1,3]
(3,5]
(5,]
RPKM of long transcript
C
100MGP-002
y = 0.4947x + 0.1156
R² = 0.3968
piRNA editing level
mRNA editing level
D
100MGP-003
A−to−C
A−to−G
A−to−T
Percent of Mutation
Editing sites
All A sites
E
100MGP-003
RPKM of piRNA
[0,0.2]
(0.2,1]
(1,3]
(3,5]
(5,]
RPKM of long transcript
F
100MGP-003
y = 0.6486x + 0.0926
R² = 0.3976
piRNA editing level
mRNA editing level
G
100MGP-004
A−to−C
A−to−G
A−to−T
Editing sites
All A sites
H
100MGP-004
RPKM of piRNA
[0,0.2]
(0.2,1]
(1,3]
(3,5]
(5,]
RPKM of long transcript
I
100MGP-004
y = 0.7742x + 0.1238
R² = 0.2831
piRNA editing level
mRNA editing level



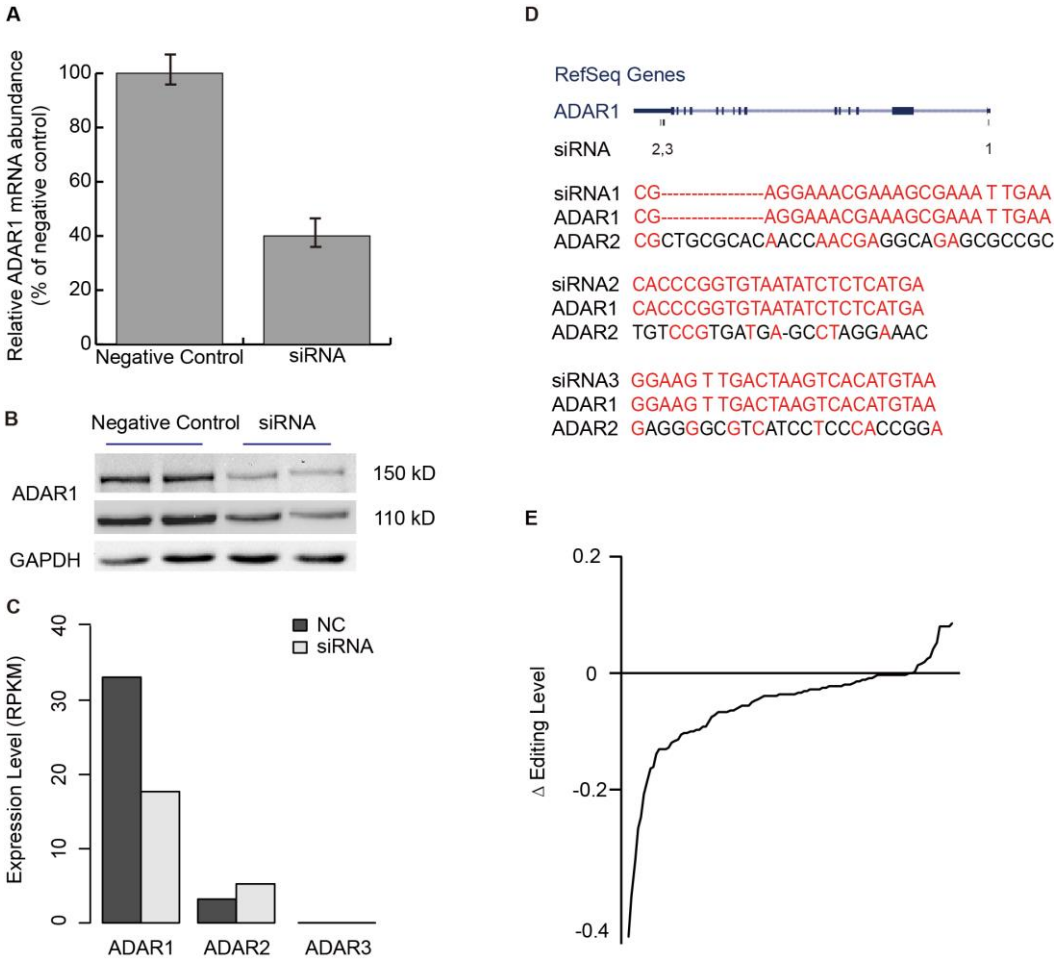
Supplementary fig. S3. Correspondence between RNA editing and piRNA regulation in three additional macaque animals and two human individuals. Correlative analyses were done on three rhesus macaque animals and two human samples. **(A**, **D, G, J and M)** The types of nucleotide substitution at the editing sites were identified according to the piRNA reads, proportions of which were summarized and shown. Distributions of the nucleotide substitutions for all "A" sites on piRNAs are also shown as a reference. **(B, E, H**, **K and N)** The expression levels of piRNAs are shown in boxplots, as binned by the expression levels of the corresponding long transcripts. **(C, F, I**, **L and O)** Editing levels were estimated from RNA-seq and small RNA-seq, and comparatively shown in dot plots.



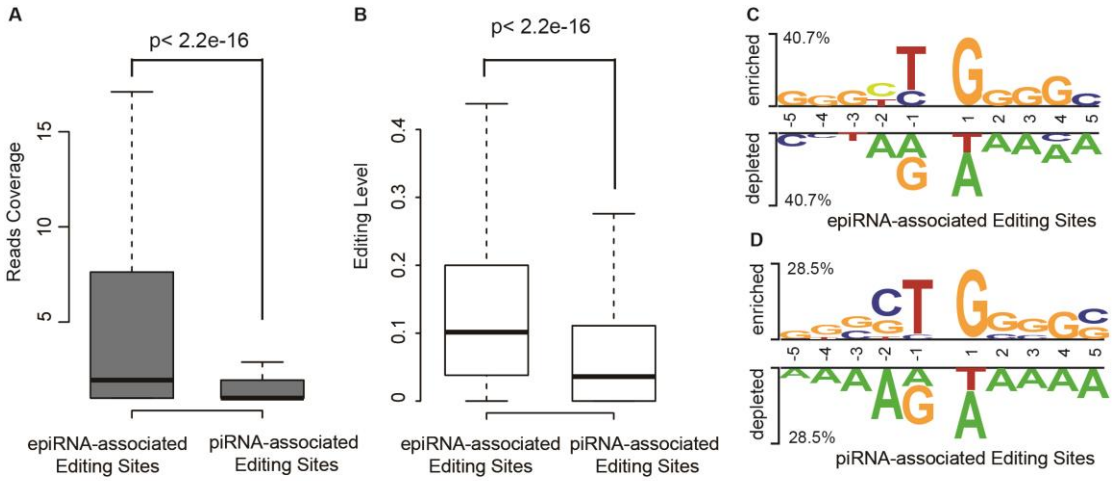
Supplementary fig. S4. Abundance of mRNA fragments in small RNA-Seq is correlated with mRNA expression in all tissues examined. For the indicated tissues, the levels of mRNA fragments in small RNA-Seq and mRNA-Seq were calculated based on RPKM. Plots for relative abundance in terms of Log_{10} transformed RPKM are shown.



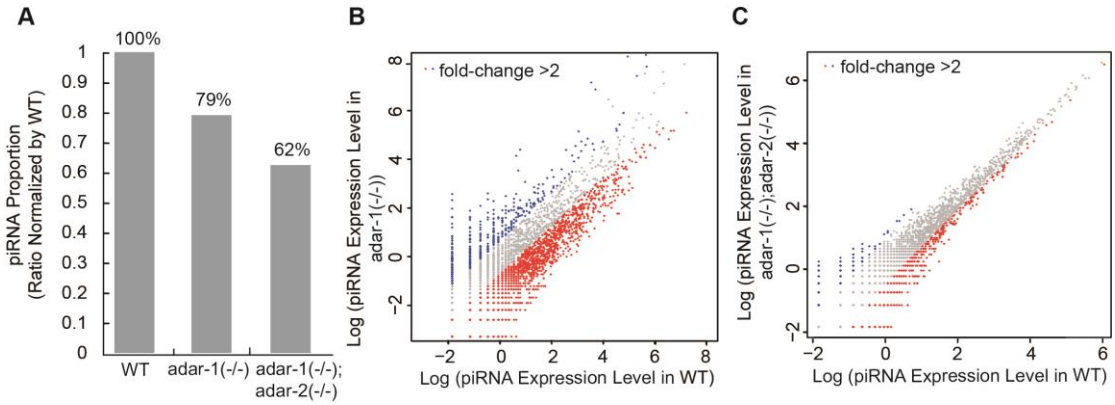
Supplementary fig. S5. Quantitative correspondence between RNA editing level and *ADAR1* expression across individuals. (**A**) The histograms summarize the distributions of Pearson correlation coefficient (*r*) values between the profiles of RNA editing level (estimated by mRNA-Seq reads) and *ADAR1* expression across four macaque animals (**Real Data**). (**B**) The distributions of *r* values between the profiles of RNA editing level (estimated by the small RNA-Seq reads) and *ADAR1* expression across four macaque animals are also summarized as histogram (**Real Data**). The *r* values estimated from 10,000-time *Monte Carlo* Simulations by randomly shuffling the profiles are also shown as Means ± SD (**Background**).



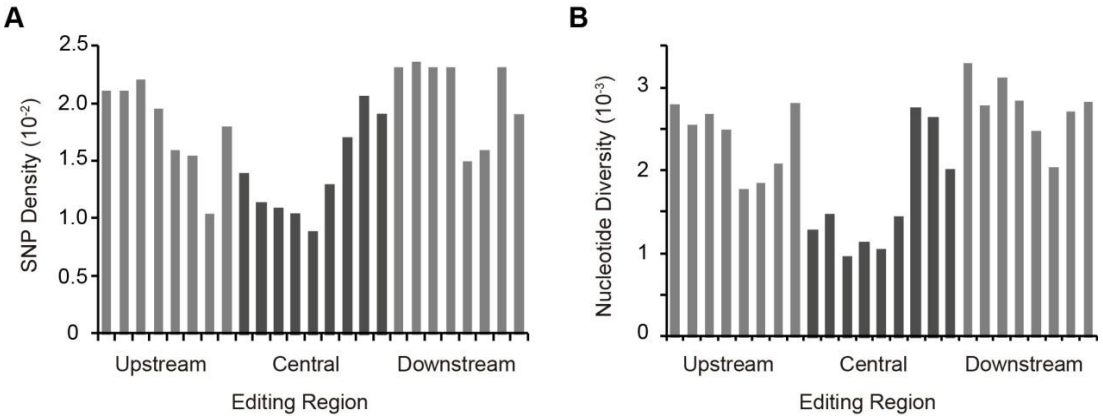
Supplementary fig. S6. Evaluation of *ADAR1* knockdown efficiency. (A) The relative *ADAR1* expression levels in the control and *ADAR1*-knockdown cells are shown in Means ± SD. (**B**) Quantification of ADAR1 proteins in the same cells as (**A**) by Western blotting. (**C**) RPKM values of *ADARs* estimated from the RNA-Seq data are shown for the control and *ADAR1*-knockdown cells. (**D**) Sequence alignments between siRNAs and the *ADARs* sequences. (**E**) RNA editing sites are plotted in ascending order according to the editing level differentials between the control and *ADAR1* knockdown samples.



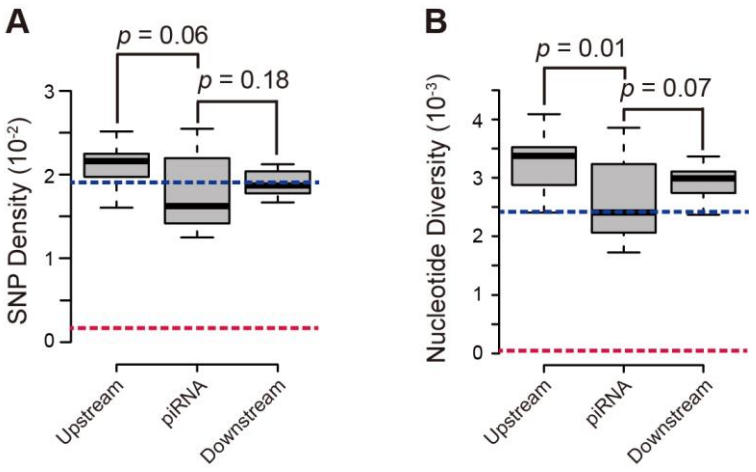
Supplementary fig. S7. **Characteristics of two classes of RNA editing sites having positional overlap with piRNAs.** The abundance of piRNAs (**A**), as well as the editing levels estimated from mRNA-Seq data (**B**) were determined for epiRNA-associated editing sites (epiRNA-associated Editing Sites) and the editing sites with no corresponding epiRNA detected in the same region (piRNA-associated Editing Sites). Degrees of correlations are summarized in boxplots. (**C**) Local sequence motifs generated by Two-Sample Logos are shown, with the extent of nucleotide preference or depletion shown in height proportional to the scale.



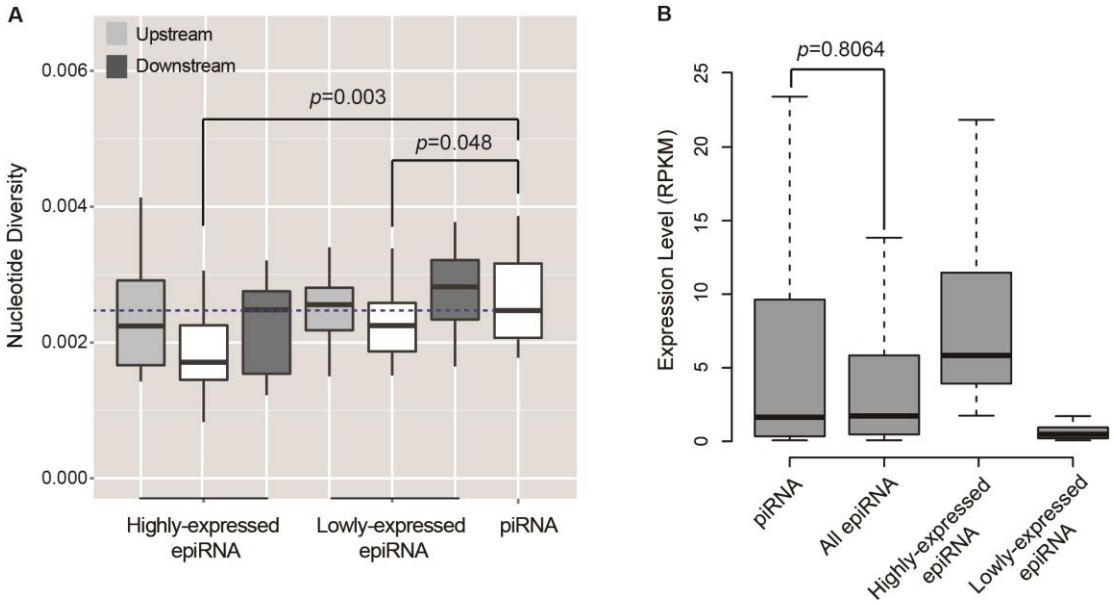
Supplementary fig. S8. **Effects of *adr* deficiency on piRNA profiles in *C. elegans*.** **(A)** The proportion of piRNA reads from each sequenced strain was calculated and further normalized by the proportion in WT. The normalized ratios are shown in the bar graph. **(B, C)** RPM of each piRNA loci in *adr* mutants vs. WT, with piRNA loci showing greater than twofold increase or decrease in expression in the mutants highlighted in blue and red, respectively.



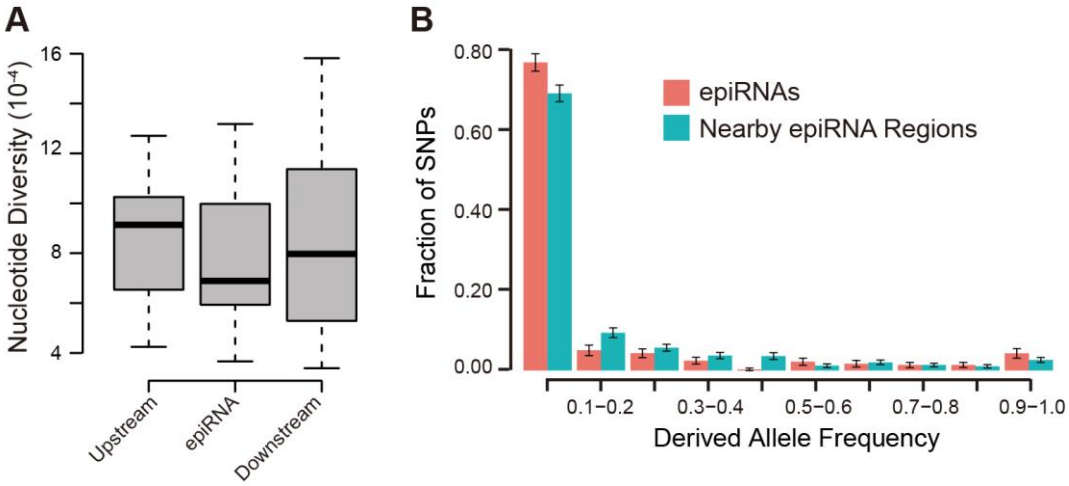
Supplementary fig. S9. Degrees of SNP density and nucleotide diversity flanking the epiRNA-associated RNA editing sites. The SNP density (**A**) and nucleotide diversity (**B**) were measured for the5' and 3' 4-bp regions flanking editing sites (Central), as well as for the distal adjacent 8-bp regions (Upstream and Downstream).



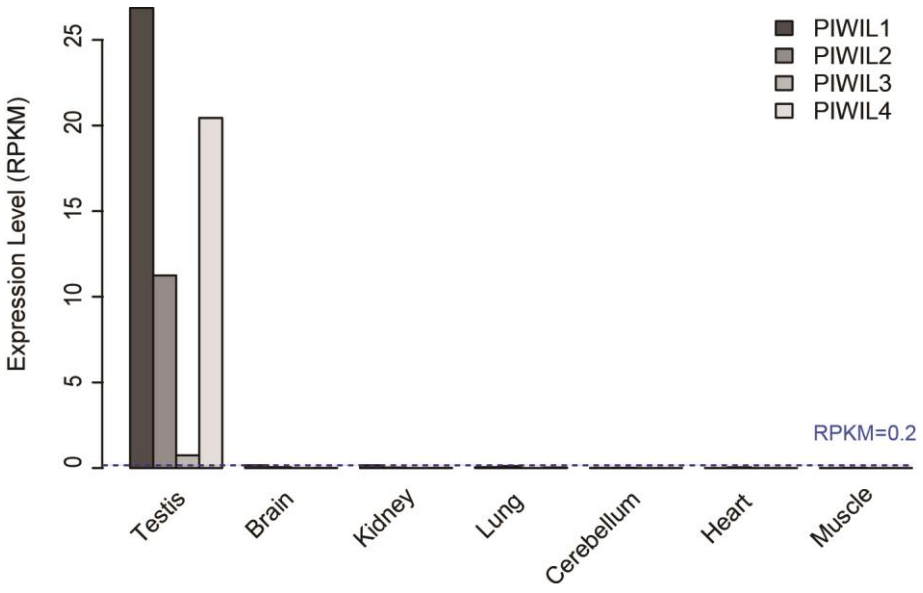
Supplementary fig. S10. Population genetics analysis of piRNA in rhesus macaque. Quantitative representation for the distribution of SNP densities (**A**), SNP nucleotide diversity (**B**), and the frequency spectra of derived allele (**C**) in macaque piRNA regions are shown, with nearby regions (Upstream and Downstream) as the references. All *p*-values were derived from *Wilcoxon* one-tail test.



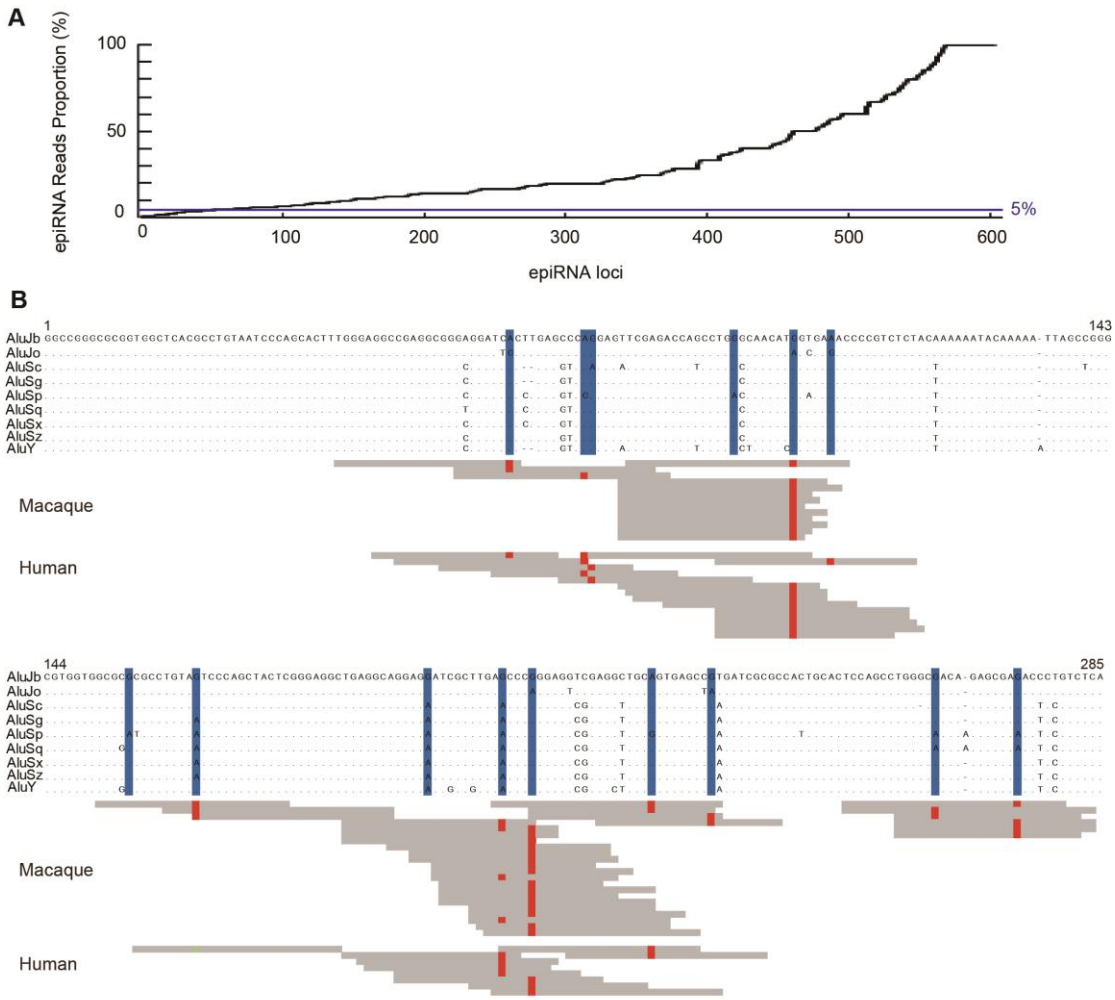
Supplementary fig. S11. Population genetics analysis of epiRNA and piRNA with different expression levels in rhesus macaque. (A) The distributions of nucleotide diversity for the canonical piRNAs and epiRNAs with high or low expression, together with their nearby regions, are shown in boxplots. The median of the nucleotide diversity for canonical piRNAs is also indicated by the dotted line as a reference. (**B**) The expression levels for canonical piRNAs and epiRNAs in (**A**) were estimated in RPKM and shown in boxplots.



Supplementary fig. S12. Population genetics analysis of human epiRNA. Quantitative representation for SNP nucleotide diversity (**A**) and the frequency spectra of derived allele (**B**) in human epiRNA regions are shown, with nearby regions (Upstream and Downstream) as the references.



Supplementary fig. S13. Expression levels of PIWIs in seven macaque tissues. The expression levels of PIWIs were calculated and shown in terms of RPKM. The background expression level was estimated as previously described (Xie, et al. 2012) and denoted by the blue dotted line.



Supplementary fig. S14.Characteristics of epiRNAs in rhesus macaque. (**A**)The proportion of epiRNA reads in each epiRNA-associated locus. (**B**) Aligned consensus sequences of selected *Alu* subfamilies are shown, with the 15 A/G discrepancies discriminating different *Alu* subfamilies highlighted in blue. epiRNAs in human and rhesus macaque were aligned accordingly, with the editing sites matching those *Alu*-associated discrepancies highlighted in red.

Supplementary table S1. Basic information of piRNA clusters in rhesus macaque and human (Data in Excel format**).**

Supplementary table S2. Basic information of epiRNAs in rhesus macaque and human (Data in Excel format**).**

Supplementary table S3. Combinational types of clustered RNA editing sites on piRNAs and long transcripts (Data in Excel format**).**

Supplementary table S4. siRNA information of ADAR1 knockdown assay (Data in Excel format**).**

Supplementary table S5. epiRNA profiles in the *C.elegans adr* mutants (Data in Excel format**).**

Supplementary table S6. Information of the 24 independent macaque animals (Data in Excel format**).**

References

Xie C, Zhang YE, Chen JY, Liu CJ, Zhou WZ, Li Y, Zhang M, Zhang R, Wei L, Li CY 2012. Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. PLoS Genet 8: e1002942. doi: 10.1371/journal.pgen.1002942