Supplementary material

A high-resolution map of functional miR-181 response elements in the thymus reveals the role of coding sequence targeting and an alternative seed match

Nikita A. Verheyden[1,*], Melina Klostermann[2,*], Mirko Brüggemann[2], Anica Scholz[3], Shady Amr[4], Chiara Lichtenthäler[5], Christian Münch[4], Tobias Schmid[3], Kathi Zarnack[2,$], and Andreas Krueger[1,$]

Supplementary Figures

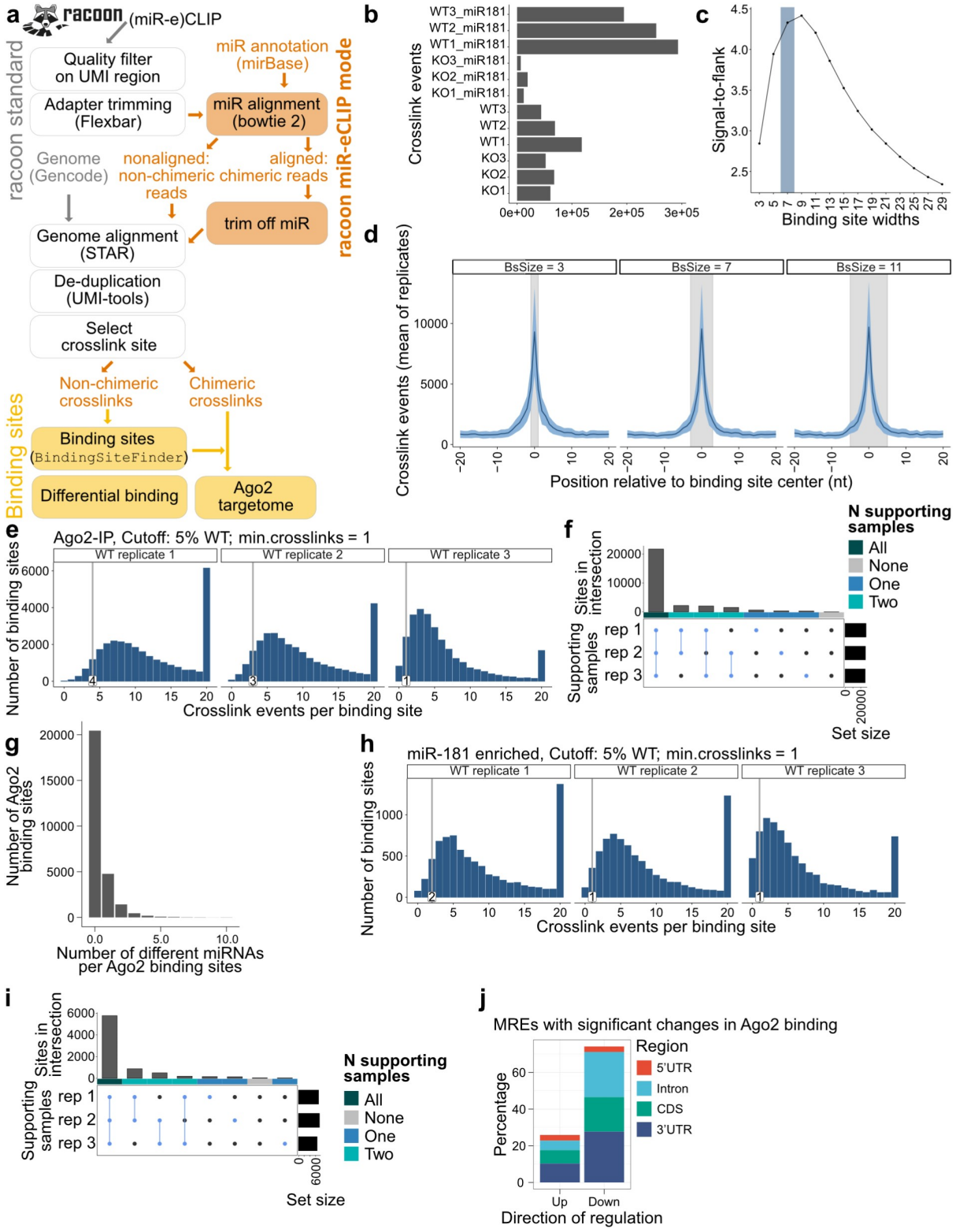

Figure S1

(a) Processing of miR-eCLIP data with `racoon` and `BindingSiteFinder`. Non-chimeric reads are processed like strander eCLIP data (grey). Chimeric miR-eCLIP reads are aligned

to the miR annotation. The chimeric part of the read is then trimmed of and the leftover read is treated like the non-chimeric reads (orange). After `racoon` processing `BindingSiteFinder` is used to define binding sites on the different subsets. These binding sites are then also further used for differential binding and together with the chimeric reads to define the AGO-targetome. **(b)** Number of crosslinks per miR-eCLIP sample. miR-eCLIP was performed on the four different conditions with three replicates each: crosslinks from Ago2-IP (WT), crosslinks from Ago2-IP (KO), and crosslinks from miR-181a/b enrichment after AGO-IP without and with miR-181 KO. **(c, d)** Determination of optimal binding site width. (c) Signal to flank ratio for binding sites from 3 to 29 nt width. (d) Distribution of crosslink events in the proximity of binding sites for binding sites of 3, 7 or 11 nt widths.
(e-h) Reproducibility filtering for Ago2 binding sites (e,f) and miR-181 MREs (g,h). (e,g) A sample-wise cutoff is defined at the lowest 5% quantile of crosslinks per binding sites. (f,h) Upset plots show the number of binding sites supported by each sample after applying the 5% cutoff. Number of supporting samples are displayed by colored bars above (dark green – all, cyan – two, blue – one, grey – none). Only binding sites supported by at least all three replicates samples are deemed reproducible. The others are filtered out. **(i)** Distribution of chimeric reads per Ago2 binding site. **(j-l)** Differential binding of Ago2 in miR-181 KO vs WT: (j) Upset plot shows the intersection of Ago2 binding sites from WT and KO conditions, which is analyzed differentially. (k) TPM cutoffs per sample. Shown are the TPMs of all binding sites ordered by their TPM (Rank). Color scale depicts the number of overlapping points. Binding sites on genes with a TPM < 50 in one of the samples (black line) are not considered for differential binding analysis. (l) The number of significantly up and down regulated binding sites are shown colored by the gene regions assigned to the binding sites (red - 5'UTR, light blue – intron, green – CDS, dark blue – 3'UTR).

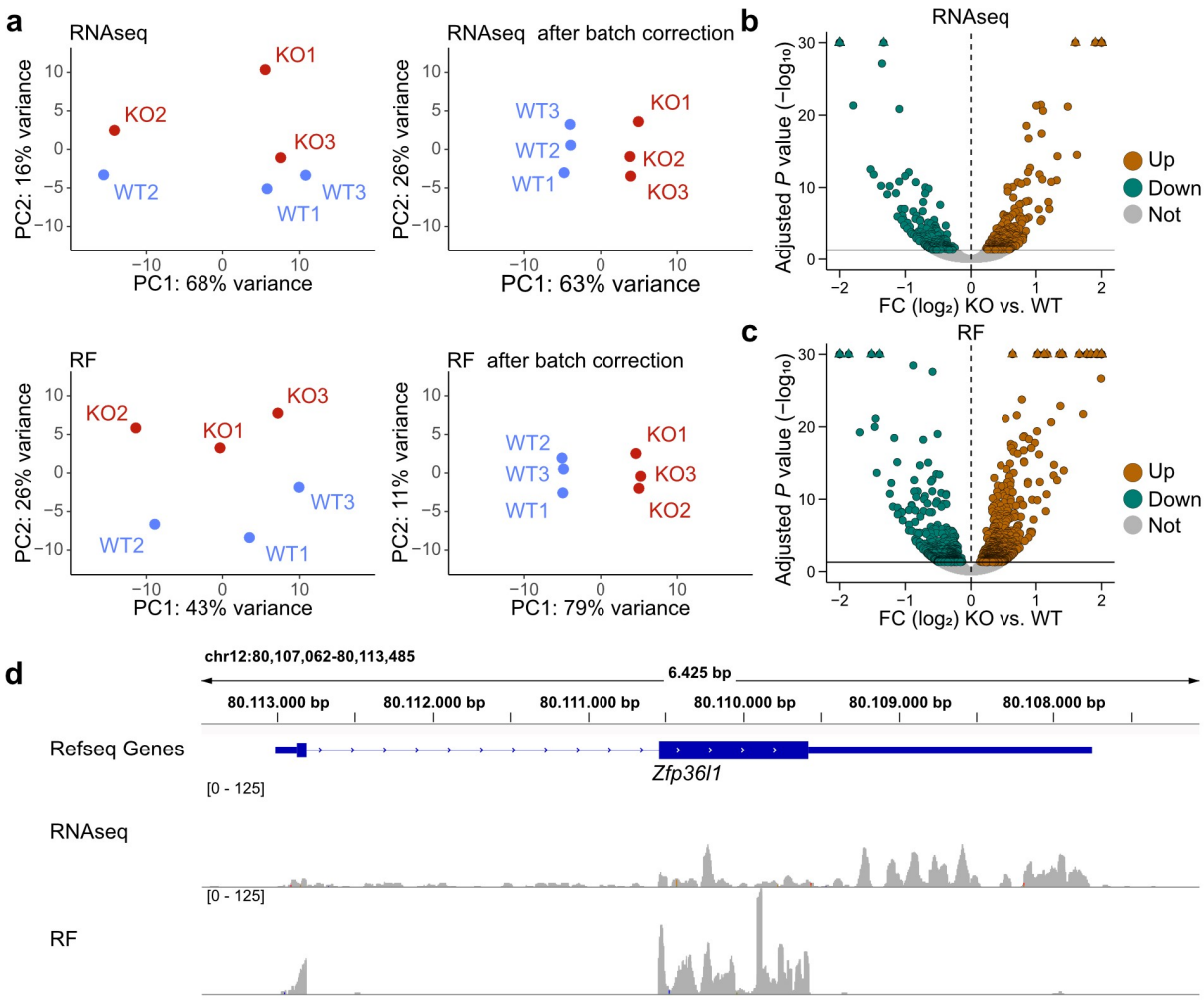

Figure S2

(a) Principal component analysis of biological replicates of RNAseq (top) and RF (bottom) data before (left) and after (right) batch correction with CombatSeq. **(b,c)** Volcano plots of RNAseq and RF data shown in Figure 3b,c plotted as $\log_2$ fold change (FC) of KO over WT vs. adjusted *P*-value ($-\log_{10}$). **(d)** Genome browser view of RNAseq (top) and RF (bottom) data in Zfp36l1.

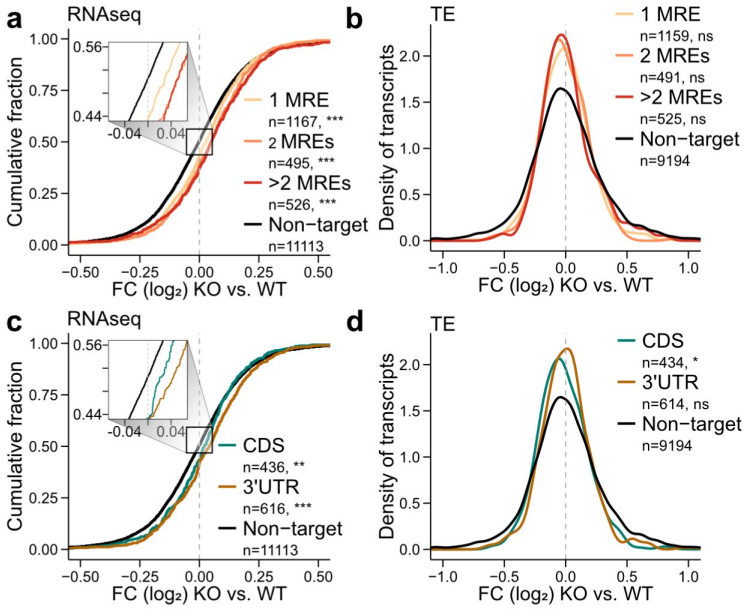

Figure S3

(a) Fold changes in transcript levels between miR-181 KO and WT conditions are shown as the cumulative density of the $\log_2$ values, for transcripts with one (yellow), two (orange), more than two (red) or no miR-181 MREs (black). Box on the left depicts zoom-in. Numbers of transcripts in each set are given. Asterisks indicate the significance of difference to genes without MREs (ns>0.05, *<0.05, **<0.01, ***<0.001, asymptotic two-sample Kolmogorov-Smirnov test). **(b)** Changes in translational efficiency (TE) for transcripts with MREs vs. non-targets shown as $\log_2$ fold-changes of KO over WT vs. density, for transcripts with one (yellow), two (orange), more than two (red) or no miR-181 MREs (black). Numbers of transcripts in each set are given (ns>0.05, *<0.05, **<0.01, ***<0.001, Welch two sample t-test). **(c)** Fold changes in transcript levels between miR-181 KO and WT conditions are shown as the cumulative density of the $\log_2$ values, for transcripts with MREs in the CDS (yellow), 3' UTR (blue), or without MREs (black). Numbers of transcripts in each set are given. Box depicts zoom-in. Only transcripts with a single MRE were used. Asterisks indicate the significance of difference to genes without MREs (ns>0.05, *<0.05, **<0.01, ***<0.001, asymptotic two-sample Kolmogorov-Smirnov test). **(d)** Changes in translational efficiency (TE) for transcripts with MREs vs. non-targets shown as $\log_2$ fold-changes of KO over WT vs. density, for transcripts with MREs in the CDS (yellow), 3' UTR (blue), or without MREs (black). Numbers of transcripts in each set are given (ns>0.05, *<0.05, **<0.01, ***<0.001, Welch two sample t-test).

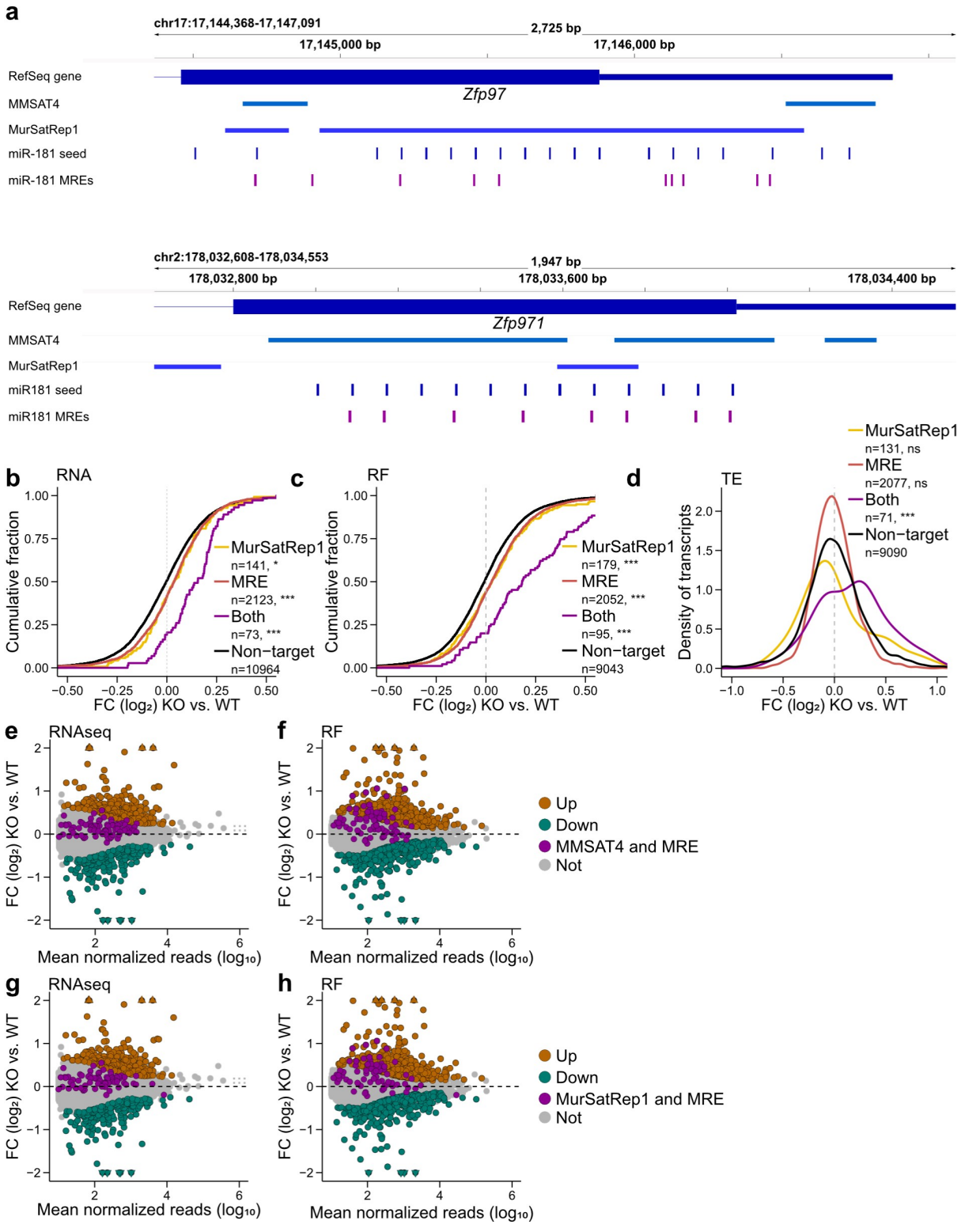
a
chr17:17,144,368-17,147,091
2,725 bp
17,145,000 bp
17,146,000 bp
RefSeq gene
Zfp97
MMSAT4
MurSatRep1
miR-181 seed
miR-181 MREs
chr2:178,032,608-178,034,553
1,947 bp
178,032,800 bp
178,033,600 bp
178,034,400 bp
RefSeq gene
Zfp971
MMSAT4
MurSatRep1
miR181 seed
miR181 MREs
b RNA
Cumulative fraction
MurSatRep1 n=141, *
MRE n=2123, ***
Both n=73, ***
Non−target n=10964
FC (log2) KO vs. WT
c RF
Cumulative fraction
MurSatRep1 n=179, ***
MRE n=2052, ***
Both n=95, ***
Non−target n=9043
FC (log2) KO vs. WT
d TE
Density of transcripts
MurSatRep1 n=131, ns
MRE n=2077, ns
Both n=71, ***
Non−target n=9090
FC (log2) KO vs. WT
e RNAseq
FC (log2) KO vs. WT
Mean normalized reads (log10)
f RF
FC (log2) KO vs. WT
Mean normalized reads (log10)
Up
Down
MMSAT4 and MRE
Not
g RNAseq
FC (log2) KO vs. WT
Mean normalized reads (log10)
h RF
FC (log2) KO vs. WT
Mean normalized reads (log10)
Up
Down
MurSatRep1 and MRE
Not

Figure S4

(a) Distribution of MMSAT4 and MurSatRep1 repeats, miR-181 seed matches, and miR-181 MREs in transcripts of two representative Zfp genes. **(b-d)** Fold changes in transcript levels (b), ribosome occupancy (c) are shown as cumulative density of the $\log_2$ values for transcripts containing MurSatRep1 repeats (yellow), transcripts containing MREs (red), transcripts containing both MREs and MurSatRep1 repeats (purple) and not containing MREs or MurSatRep1 repeats (black); Asterisks indicate the significance of difference to transcripts without MREs (ns>0.05, *<0.05, **<0.01, ***<0.001, asymptotic two-sample Kolmogorov-Smirnov test). (d) translational efficiency is shown as density (ns>0.05, *<0.05, **<0.01, ***<0.001, Welch two sample t-test). **(e-h)** Transcripts with significant differences (adjusted *P* value < 0.05, Benjamini-Hochberg correction) in RNA abundance (RNAseq) (e,g) or ribosome occupancy (RF) (f,h) are plotted as log2 fold change (FC) of KO over WT vs. mean normalized read count as shown in Figure 3b,c. transcripts containing both MREs and MMSAT4 repeats (e,f) or MREs and MurSatRep1 repeats (g,h) are highlighted in purple.

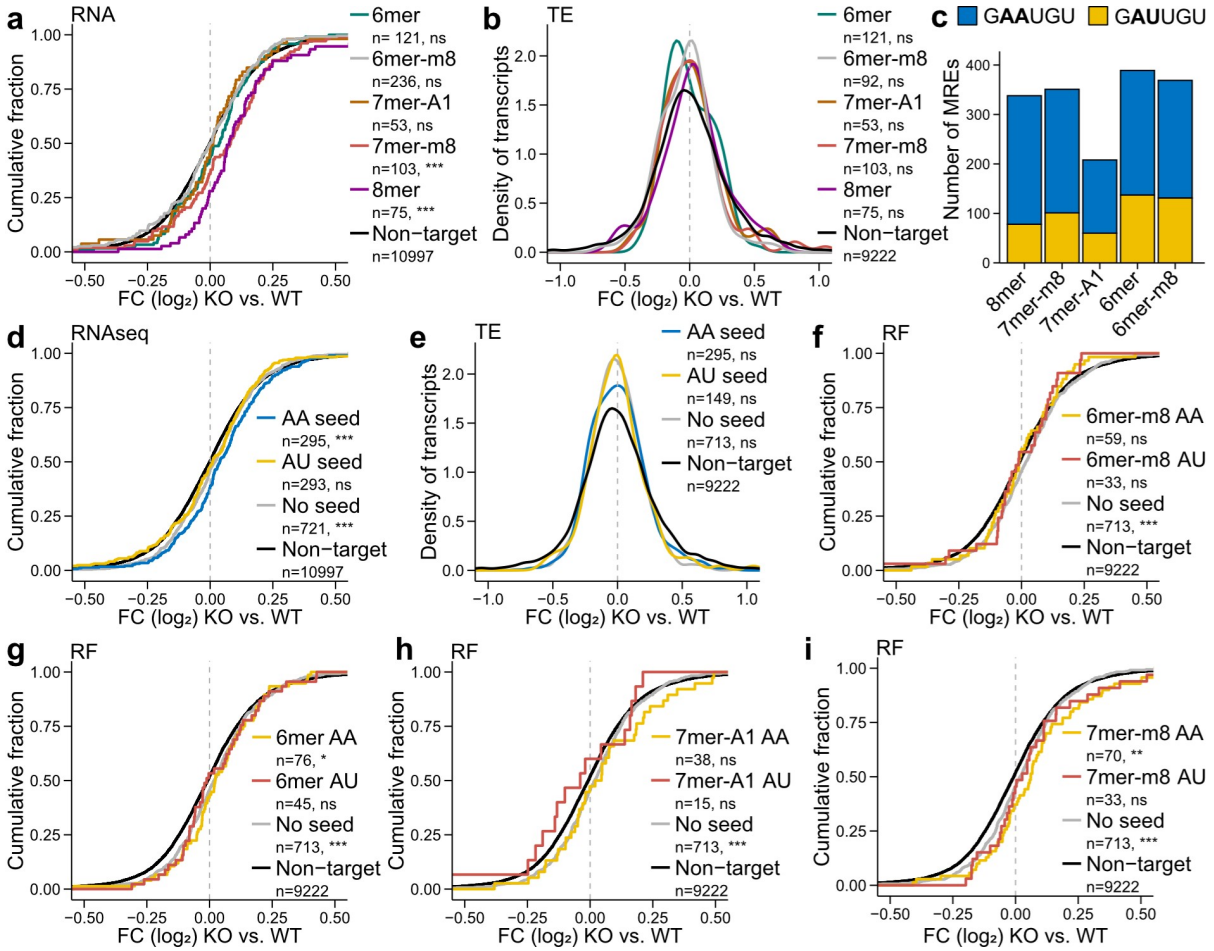

Figure S5

(a) Fold changes in transcript levels between miR-181 KO and WT conditions are shown as the cumulative density of the $\log_2$ values. Comparison of transcripts with MREs containing 6mer (green), 6mer-m8 (grey), 7mer-A1 (brown), 7mer-m8 (red) and 8mer seed matches (purple). Only transcripts with a single MRE were used. Numbers of transcripts in each set are given. Asterisks indicate the significance of difference to genes without MREs (ns>0.05, *<0.05, **<0.01, ***<0.001, asymptotic two-sample Kolmogorov-Smirnov test). **(b)** Changes in translational efficiency (TE) for transcripts with MREs vs. non-targets shown as $\log_2$ fold-changes of KO over WT vs. density. Comparison of transcripts with MREs containing 6mer (green), 6mer-m8 (grey), 7mer-A1 (brown), 7mer-m8 (red) and 8mer seed matches (purple). Only transcripts with a single MRE were used. Numbers of transcripts in each set are given (ns>0.05, *<0.05, **<0.01, ***<0.001, Welch two sample t-test). **(d)** Fold changes in transcript levels between miR-181 KO and WT conditions are shown as the cumulative density of the $\log_2$ values. Comparison of transcripts with MREs with the canonical AA seed match sequence (blue), the alternative AU seed match (yellow), MREs without seed match (grey) or genes without MREs (black). Only transcripts with a single MRE were used. Numbers of transcripts in each set are given. Asterisks indicate the significance of difference to genes without MREs (ns>0.05, *<0.05, **<0.01, ***<0.001, asymptotic two-sample Kolmogorov-Smirnov test). **(e)** Changes in translational efficiency (TE) for transcripts with MREs vs. non-targets shown as $\log_2$ fold-changes of KO over WT vs. density. Comparison of transcripts with MREs with the canonical AA seed match sequence (blue), the alternative AU seed match (yellow), MREs without seed match (grey) or genes without MREs (black). Only transcripts with a single MRE were used. Numbers of transcripts in each set are given (ns>0.05, *<0.05, **<0.01, ***<0.001, Welch two sample t-test). (f-i) Comparison of transcripts with MREs with the canonical AA seed match sequence (yellow), the alternative AU seed match (red) for 6mer-m8 (f), 6mer (g), 7mer-A1 (h) and 7mer-m8 (i), MREs without seed match (grey) or genes without MREs (black).

Only transcripts with a single MRE were used. (ns>0.05, *<0.05, **<0.01, ***<0.001, asymptotic two-sample Kolmogorov-Smirnov test).

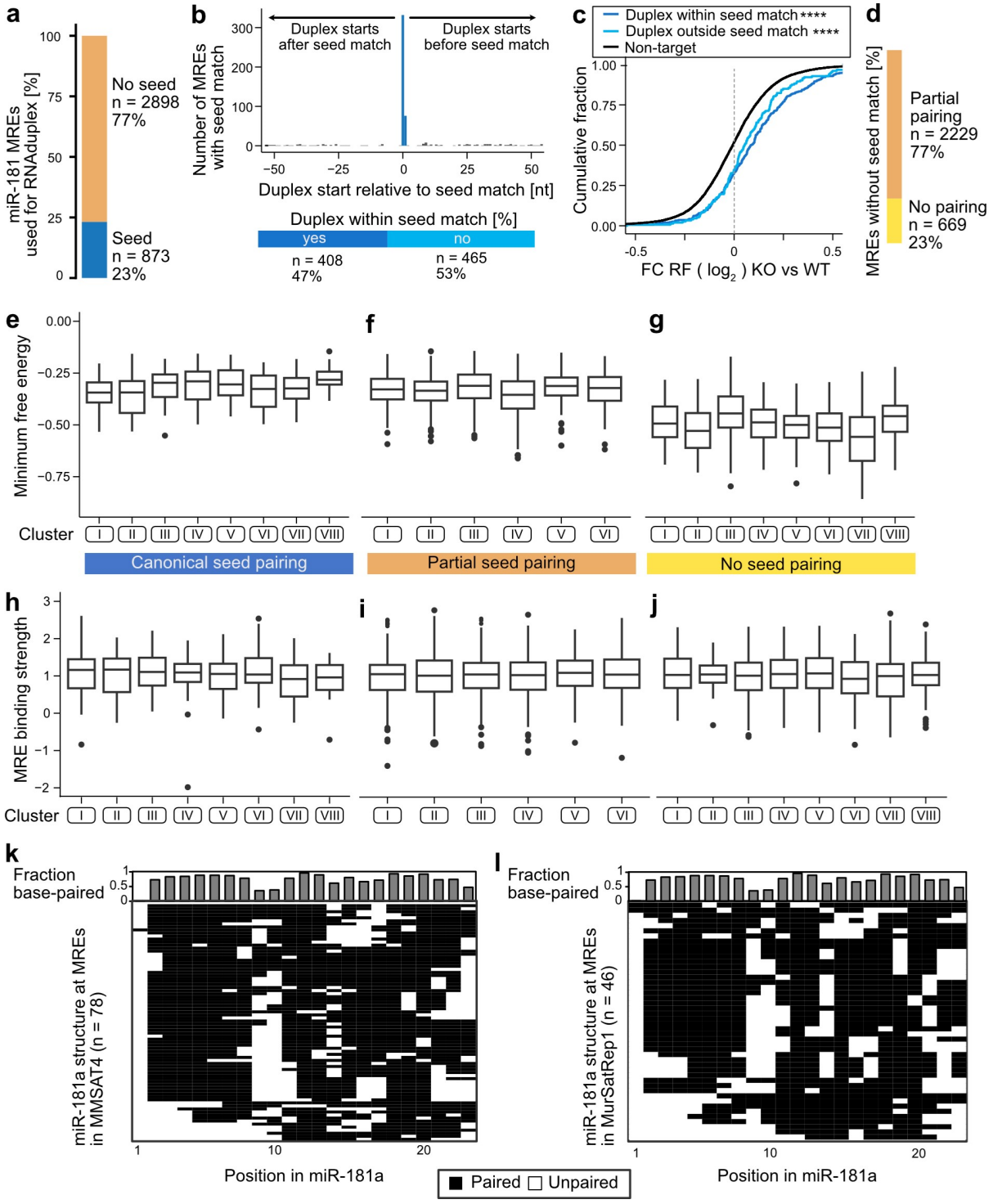

Figure S6

(a) Ratio of miR-181 MREs used for RNAduplex containing a seed match (blue) or not (orange). **(b)** Top: Histogram of duplex start position relative to the start of the seed match in MRE sequences with a seed match. Shown is a window spanning 50 nt before and 50 nt after the start of the seed match. Bottom: Ratio of duplexes starting at the start of the seed match or one nt before (dark blue) to duplexes starting at other positions (light blue). **(c)** Fold changes in ribosome occupancy between miR-181 KO and WT conditions are shown as the cumulative density of the $\log_2$ values. Shown are duplexes starting at the start of the seed match or one nt before (dark blue), duplexes starting at other positions (light blue) and genes without MREs (black). Asterisks indicate the significance of difference to genes without MREs (**** - $p < 0.001$, Kolmogorov Smirnov test). **(d)** Ratio of miR-181 MREs without a seed match with either

partial (orange) or no (yellow) pairing within the first 8 nts of the miR-181a sequence. **(e-g)** Boxplots of the normalized minimum free energy per cluster for canonical duplex formation (e), partial seed pairing (f), or no seed pairing (g). Center line depicts the median, box encloses the 25% and 75% quantiles. Whisker is the maximum value of the data that is within 1.5 times the interquartile range over or under the 75th (upper) or 25% (lower) percentile respectively. **(h,i,j)** Boxplots of the MRE binding strength per cluster for canonical duplex formation (h), partial seed pairing (i), or no seed pairing (j). Center line depicts the median, box encloses the 25% and 75% quantiles. Whisker is the maximum value of the data that is within 1.5 times the interquartile range over or under the 75th (upper) or 25% (lower) percentile respectively. **(k,l)** Heatmap of miR-181a structures in duplexes with targets overlapping MMSAT4 (k) or MurSatRep1 (l) repeats. White – unpaired, black paired. Bars on the top annotate the paired ratio per nucleotide.