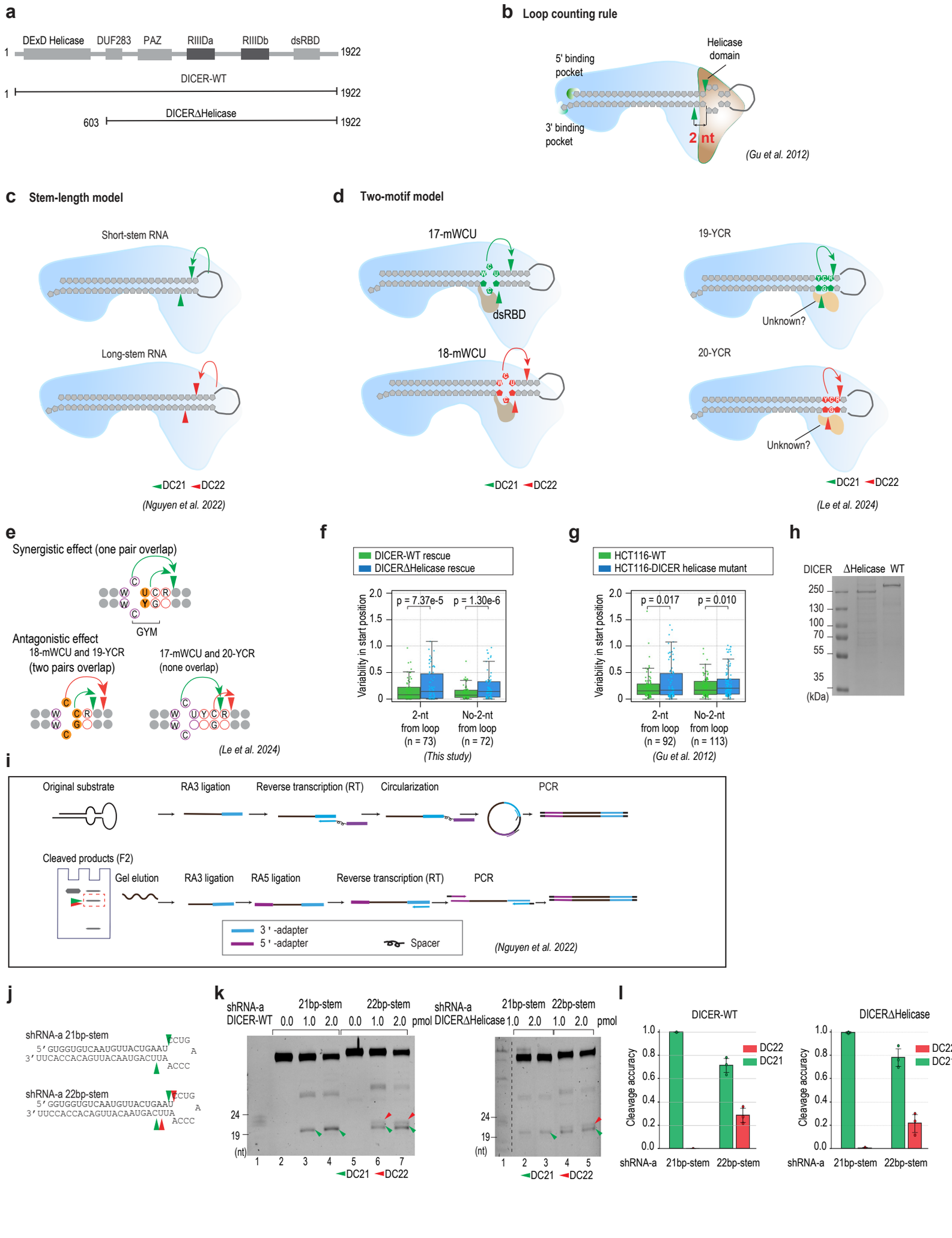
Supplementary Information

Determinants of selectivity in the dicing mechanism

Thi Nhu-Y Le, Cong Truc Le, and Tuan Anh Nguyen

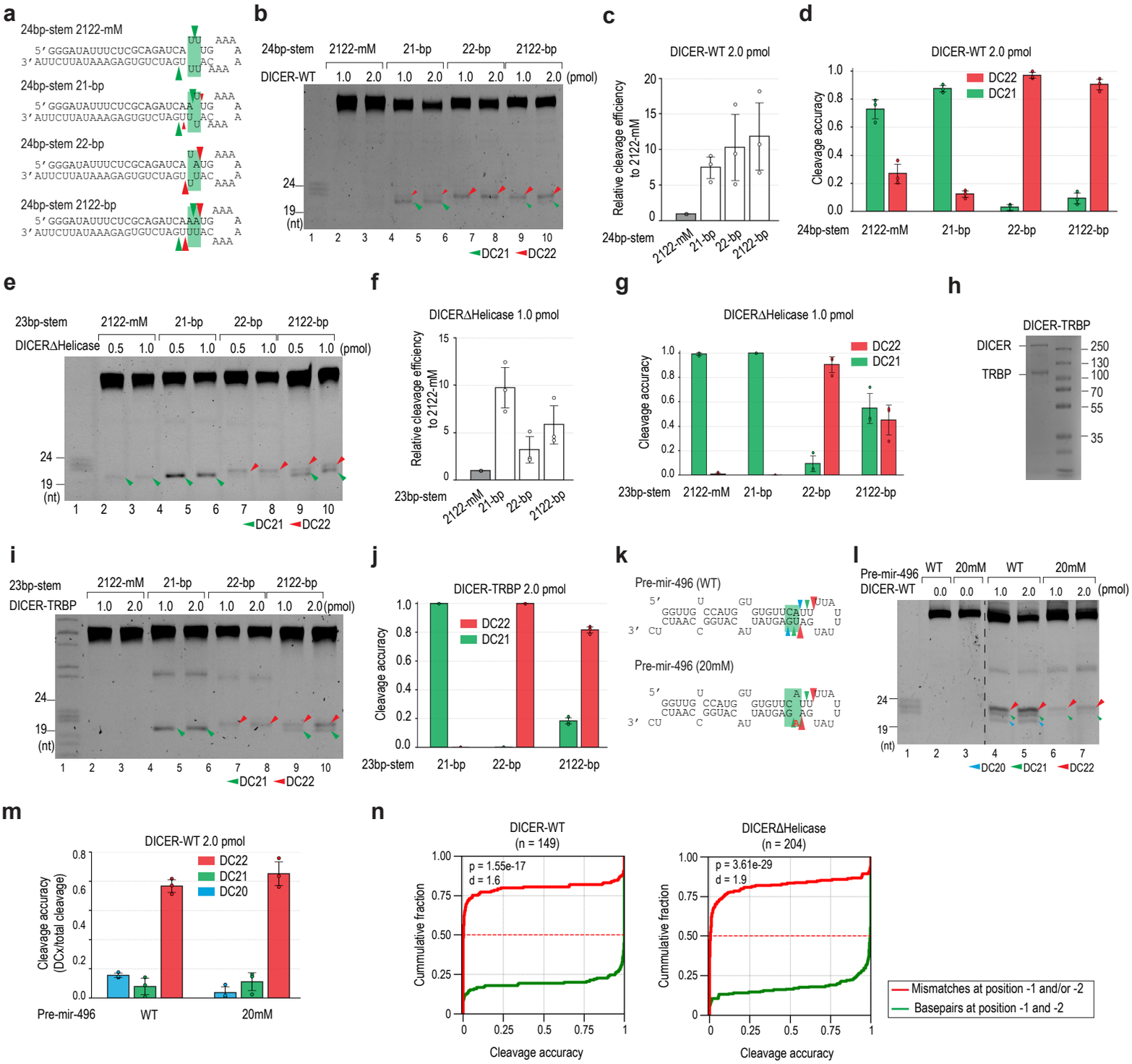
Supplementary Fig. 1



Supplementary Fig. 1 Reevaluating the role of helicase in the loop-counting rule

a, DICER domain structure.The schematic illustration of the domain architecture of DICER, with the numbers indicating the initial and terminal residues of the protein construct employed in this study. **b,** Loop counting rule. This existing model emphasizes the significance of the helicase domain in measuring the 2 nucleotide (nt) distance between the RNA loop and the DICER cleavage site. **c,** Stem length model. This model describes the cleavage site preferences, with DC22 favored by long-stemmed RNAs at 22 nt from the 5p-end, and DC21 favored by short-stemmed RNAs at 21 nt from the 5p-end. **d,** Two-motif model. The dsRBD domain interacts with mWCU motifs to determine the cleavage site, with the specific mechanism for YCR recognition remaining elusive. mWCU motifs situated 17 and 18 nucleotides from the 5p-end of RNAs promote DICER cleavage at DC21 and DC22, respectively. mWCUs are characterized by a mismatch at the "C" position. The complete catalogue of mWCU motifs is documented in Le et al., 2024. YCR motifs, located 19 and 20 nucleotides from the 5p-end, also stimulate cleavage at DC21 and DC22, respectively. YCRs feature a C-G or G-C pair at the "C" position and A-U or G-U pairs at the "R" positions, with the full list of YCR motifs available in Le et al., 2024. **e,** Synergistic and antagonistic interactions of mWCU and YCR. The combined effect of 17-mWCU with 19-YCR or 18-mWCU with 20-YCR coordinates to stimulate cleavage at DC21 or DC22. Conversely, the combination of 18-mWCU with 19-YCR may have an antagonistic effect. **f, g** Variability in 3p-miRNAs start position analysis. A comparative study of the variability in the start position of 3p-miRNAs originating from '2-nt from loop' and 'no-2-nt from loop' pre-miRNAs in DICER knockout (KO) cells rescued with either DICER-WT or DICERΔHelicase **(f)**, as well as in wild-type versus DICER mutant HCT116 cells **(g)**. The variability value reflects the degree to which DICER cleavage accuracy is altered when processing the pre-miRNA that produces that specific miRNA. Each dot represents one pre-miRNA. Wilcoxon signed rank tests were used to examine statistical change. n indicated the sample size. The median value (center line), lower quartile and upper quartile (box edges), and whiskers extended to ±1.5 × IQR (interquartile range) were indicated in each boxplot. The detailed methodology for this analysis is outlined in the Methods section. **h,** The purified DICERΔHelicase and DICER-WT proteins analyzed by SDS-PAGE. **i,** The cloning scheme was applied for the original substrate (OS) and cleavage product (CP or F2). The detail of oligos and enzymes used in this scheme are presented in the Methods section. **j,** Representation of shRNA structures and sequences, with green and red arrowheads indicating the DC21 and DC22 cleavage sites, respectively. **k,** In vitro DICER cleavage assays performed on RNAs depicted in j using DICER-WT and DICERΔHelicase. **l,** Assessment of cleavage accuracy at DC21 and DC22 for both DICER-WT and DICERΔHelicase based on data from three duplicate assays as shown in **(k)**. Data values are shown as dots, and the error bars are presented with 95%

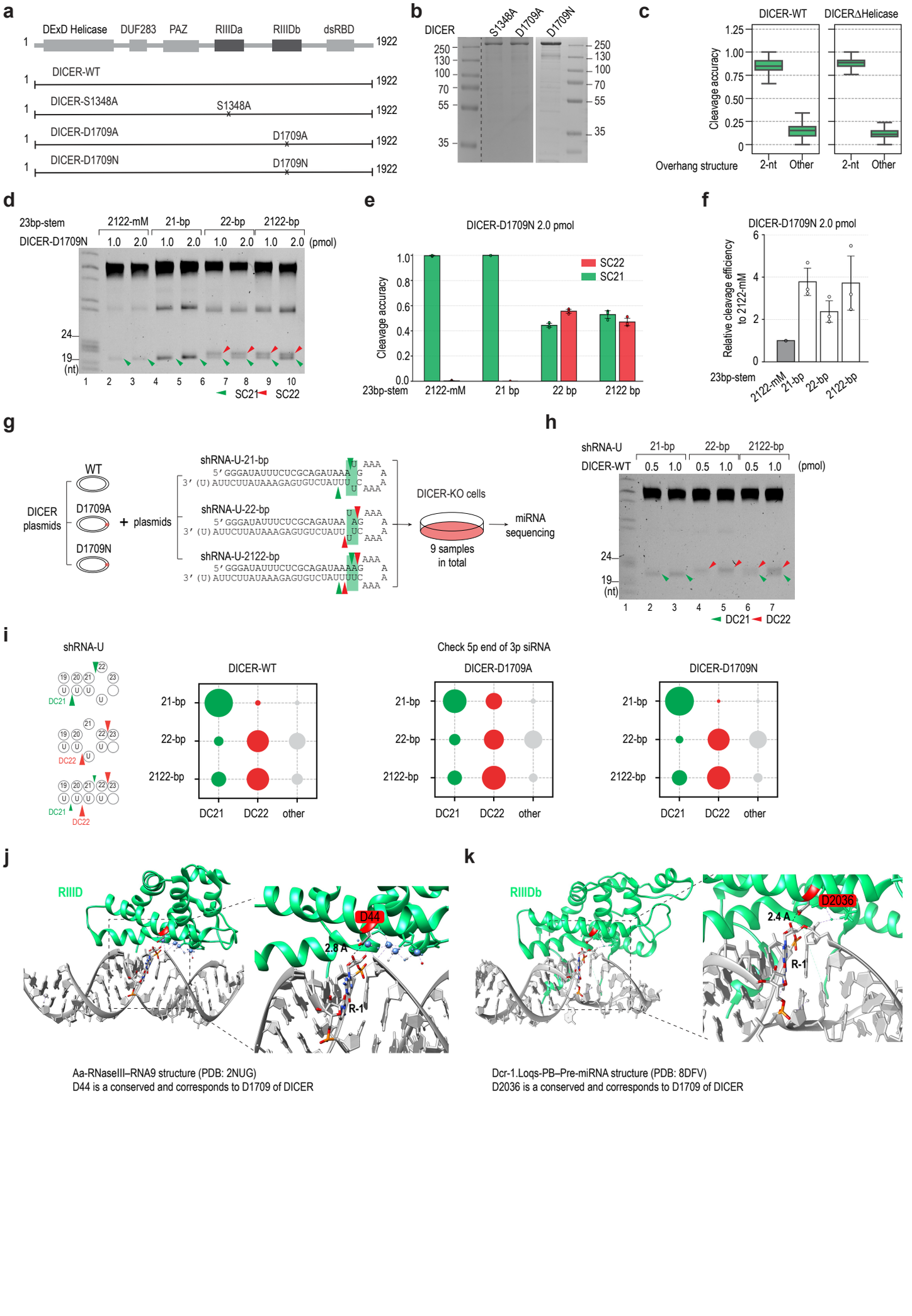
Supplementary Fig. 2



Supplementary Fig. 2 The bipartite pairing rule

a, Representation of shRNA structures and sequences. **b,** In vitro DICER cleavage assays performed on RNAs depicted in **(a)** using DICER-WT. **c,** Calculation of DICER-WT cleavage efficiency on various shRNAs from data obtained in three replicate assays as depicted in **(b)** normalized to 2122-mM RNA. **d,** Evaluation of cleavage accuracy at DC21 and DC22 for DICER-WT based on three repetitive assays as shown in **(b). e,** In vitro DICER cleavage assays performed on RNAs depicted in Fig. 2d using DICERΔHelicase. **f,** Calculation of DICERΔHelicase cleavage efficiency on various shRNAs from data obtained in three replicate assays as depicted in **(e)** normalized to 2122-mM RNA. **g,** Evaluation of cleavage accuracy at DC21 and DC22 for DICERΔHelicase based on three repetitive assays as shown in **(e). h,** The purity of the DICER-TRBP was displayed on SDS-PAGE. **i,** In vitro DICER cleavage assays performed on RNAs depicted in **Fig. 2d** using DICER-TRBP. **j,** Evaluation of cleavage accuracy at DC21 and DC22 for DICERΔHelicase based on three repetitive assays as shown in **(i). k,** Representation of pre-mir-496 structures and sequences, with blue, green and red arrowheads indicating the DC20, DC21 and DC22 cleavage sites, respectively. **l,** In vitro DICER cleavage assays performed on RNAs depicted in **(k)** using DICER-WT. **m,** Evaluation of cleavage accuracy at DC20, DC21 and DC22 for DICER-WT based on three repetitive assays as shown in **(l). n,** Accuracy of cellular DICER cleavages measured by the level of isomiRs. miRNAs originated from the 3p-strand of pre-miRNAs were utilized in this analysis. The raw read count was normalized to obtain RPM value (reads per million). Each isomiR was assigned to one of the two groups depending on the presence of the mismatch at position -1 and -2 related to expected cleavage site on the 5p-strand. For each pre-miRNA, the accuracy of cleavage for each group of isomiRs was measured by ∑RPMisomiR, group/∑RPMisomiR, pre-miRNA, in which ∑RPMisomiR, group and ∑RPMisomiR, pre-miRNA are the sum of RPM of all isomiRs of the group and of the pre-miRNA, respectively. p-value (p) obtained from paired samples t-tests and Cohen's d value (d) were indicated. n indicated the number of pri-miRNAs detected in the sample. The red dash line indicated the median line. In the bar plot representing quantitative data from the gel, data values are shown as dots, and the error bars are presented with 95% confidence intervals.

Supplementary Fig. 3



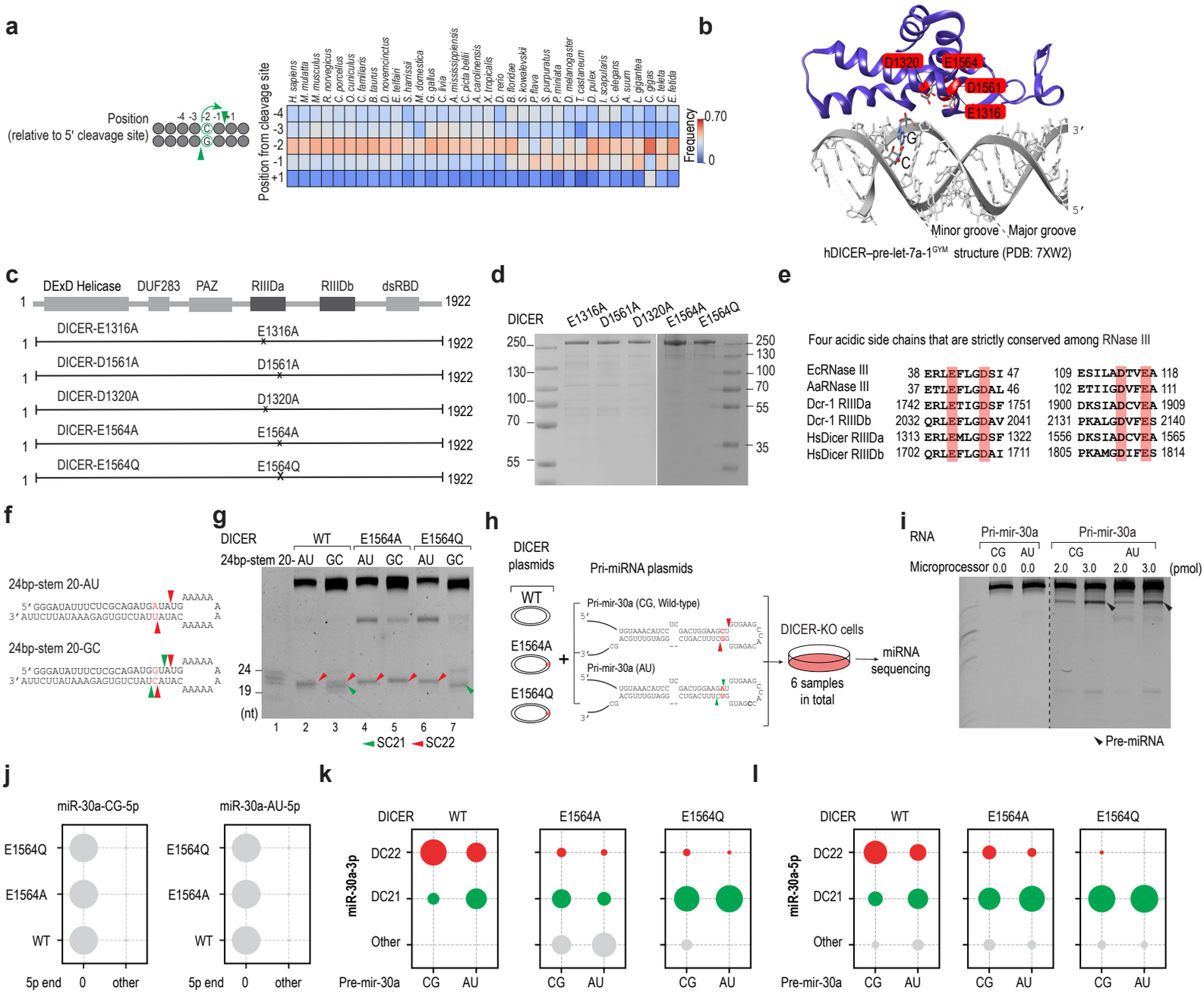
Aa-RNaseIII–RNA9 structure (PDB: 2NUG)
D44 is a conserved and corresponds to D1709 of DICER

Dcr-1.Loqs-PB–Pre-miRNA structure (PDB: 8DFV)
D2036 is a conserved and corresponds to D1709 of DICER

Supplementary Fig. 3 Molecular basis of the bipartite pairing rule

a, DICER constructs and mutations. Mutation sites on each construct are marked with asterisks. **b,** The purity of the various DICER mutants was displayed on SDS-PAGE. **c,** Accuracy of the processing events by DICER or DICERΔ Helicase on randomized shRNAs at the positions that generate cleaved fragments with or without a 2-nt overhang structure. All shRNA variants in the high-throughput data harboring symmetric structures (no bulges in the stem) were utilized in this analysis. For each shRNA variants, the accuracy score of each cleavage group ("2-nt overhang" cleavage or "Other" cleavage) was calculated by ∑RPMgroup/∑RPMshRNA, in which ∑RPMgroup and ∑RPMshRNA are the sum of RPM of all cleaved fragments of the group and of the variant, respectively. **d,** In vitro DICER cleavage assays performed on RNAs depicted in **Fig. 2d** with DICER-D1709N. **e,** Evaluation of cleavage accuracy at DC21 and DC22 for DICER-D1709N based on three repetitive assays as shown in **(d). f,** Calculation of DICER-D1709N cleavage efficiency on various shRNAs from data obtained in three replicate assays as depicted in **(d). g,** DICER-WT or mutants were co-transfected with shRNA plasmids in HCT116 DICER-KO cells, each sample was conducted two times. Small RNA sequencing was performed. **h,** In vitro DICER cleavage assays performed on shRNA-U depicted in **(g)** using DICER-WT. **i,** The accuracy of cellular DICER cleavage at sites DC21 and DC22 was assessed for various shRNAs. **j,** Aa-RNase III - RNA interaction. A schematic representation based on the PDB entry 2NUG illustrates the Aa-RNase III enzyme in a complex with RNA. D44 forms a hydrogen bond with the RNA base at position -1, depicted by sticks for the molecular structures and a dashed line for the hydrogen bond. D44 is a conserved residue, equivalent to D1709 in DICER. **k,** Dcr-1.Loqs-PB-Pre-miRNA complex. This illustration, derived from PDB entry 8DFV, shows Dcr-1.Loqs-PB bound to a pre-miRNA. The hydrogen bond between D2036 of Dcr-1 and the base at position -1 of the pre-miRNA is shown with sticks and a dashed line. D2036, a conserved residue, aligns with D1709 of DICER in terms of functional analogy. In the bar plot representing quantitative data from the gel, data values are shown as dots, and the error bars are presented with 95% confidence intervals.

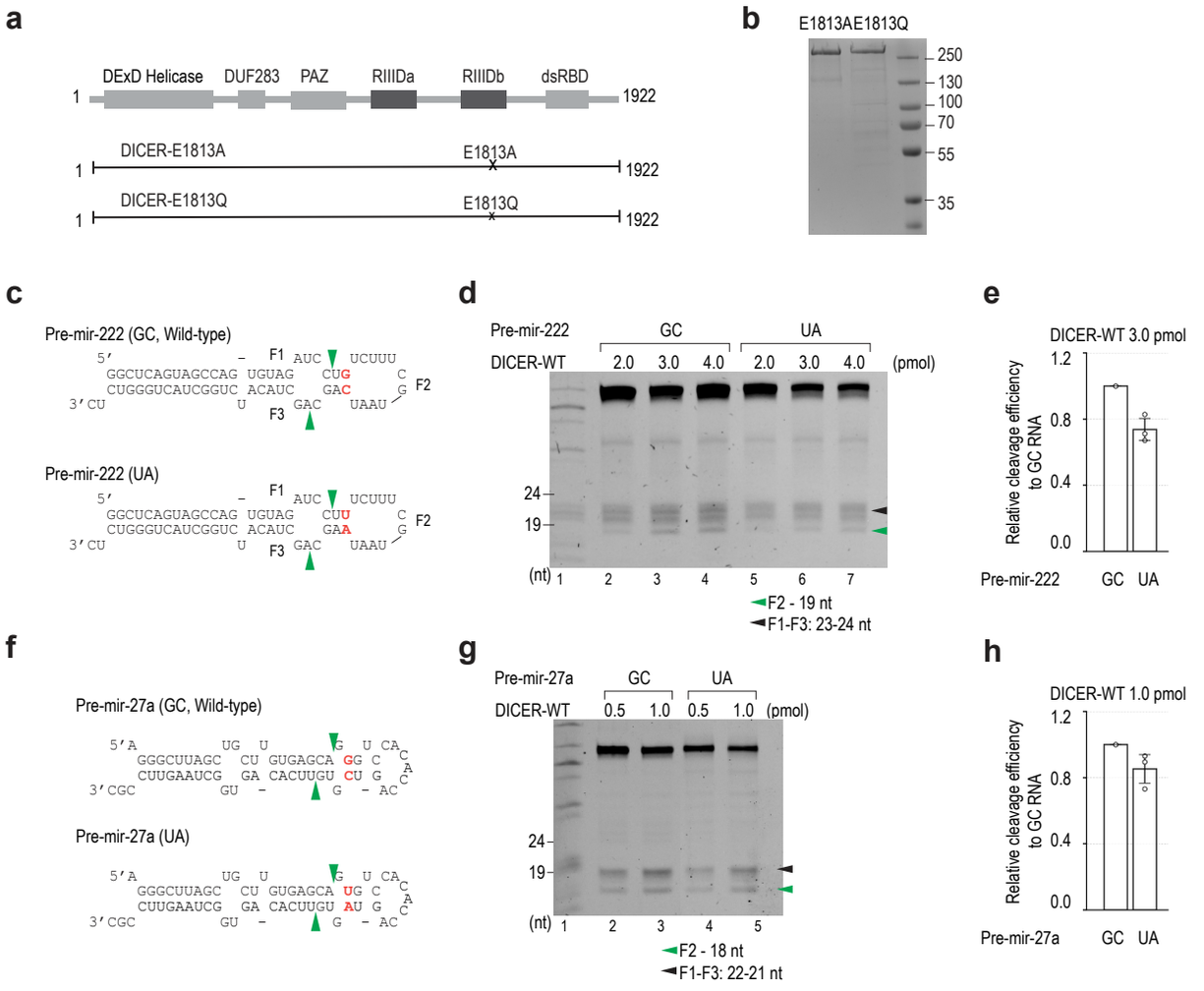
Supplementary Fig. 4



Supplementary Fig. 4 YCR recognizing mechanism

a, C-G and G-C pair enrichment in pre-miRNAs. The frequency of C-G and G-C base pairs within pre-miRNAs from various animal species is quantified. The enrichment is calculated as the ratio of pre-miRNAs containing C-G or G-C pairs at a specific position to the total number of pre-miRNAs analyzed. **b,** DICER/RNA complex illustration. A schematic based on the PDB entry 7XW2 depicts the DICER/RNA complex at the dicing stage. In the illustration, four acidic residues crucial for the mechanism are highli ghted in red, and the C-G pair in the YCR motif is color-coded for clarity. **c,** DICER protein constructs and mutants. Mutation sites are indicated with asterisks, and the specific mutations utilized in this study are listed. **d,** Mutant DICER enzyme purity was examined by SDS-PAGE. **e,** RIIID domain sequence alignment. Sequence alignments of RIIID domains from various species are presented. These sequences were retrieved from UniProt (https://www.uniprot.org/). The alignment highlights that four acidic residues are highly conserved across the catalytic centers of these enzymes. **f,** Diagrams of shRNA molecules illustrate their structures and sequences. Green and red arrowheads mark the DC21 and DC22 cleavage sites, respectively. **g,** In vitro DICER cleavage assays for RNAs described in **(f). h.** Wild-type DICER or mutants were co-transfected with pri-mir-30a plasmids in HCT116 DICER-KO cells, followed by small RNA sequencing. **I.** In vitro pri-miRNA cleavage assays were conducted using the Microprocessor complex on the RNAs described in **(h). j.** The percentage of miR-30a-5p containing the first nucleotide mapped to the canonical DROSHA cleavage site is indicated as 0. **k, l.** The accuracy of cellular cleavage by DICER-WT or mutants was assessed by analyzing isomiRs of miR-30a-3p and miR-30a-5p from the transfected cells in **(h)**. The DICER cleavage site was determined using the 5p-end nucleotide of miR-30a-3p **(k)** and the 3p-end nucleotide of miR-30a-5p **(l)** the data was obtained from two biological replicate experiments.

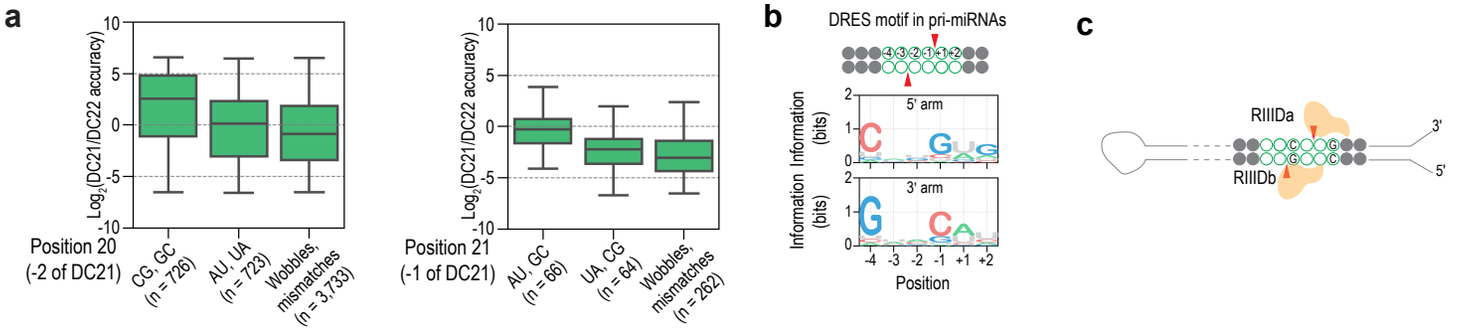
Supplementary Fig. 5



Supplementary Fig. 5 Discovery of the secondary YCR motif

a, DICER protein constructs and mutants. Mutation sites are indicated with asterisks, and the specific mutations utilized in this study are listed. **b,** Mutant DICER enzyme purity was examined by SDS-PAGE. **c, f** Diagrams of pre-mir-222 in **(c)** and pre-mir-27a in **(f)** illustrate their respective structures and sequences. **d, g** In vitro DICER cleavage assays for RNAs described in (**c** and **f**). **e, h** Bar graphs display the relative cleavage efficiency for three repeated assays in **(d)** and **(g)**, normalized to wild-type pre-miRNAs. In the bar plot representing quantitative data from the gel, data values are shown as dots, and the error bars are presented with 95% confidence intervals.

Supplementary Fig. 6



Supplementary Fig. 6 Multiple mechanisms to ensure the accuracy of dicing

a, Ratio of cleavage accuracy at DC21/DC22 (log_2) for shRNA variants containing various nucleotide combinations at positions -1 and -2, relative to the cleavage at DC21. n indicated the number of shRNA variants. **b,** C-G and G-C pair enrichment in DRES (DROSHA dsRNA recognition sites) motifs. Enrichment analysis was conducted on previously identified DRES motifs to determine the frequency of C-G and G-C pairs. This analysis utilized the Logomaker software package (Tareen and Kinney, 2020). In the graphical output, the height of each nucleotide stack is proportional to the information content at that position in the sequence. For a detailed explanation of the analysis, please refer to the Methods section. **c,** Schematic representation of DRES positioning. The scheme illustrates the location of DRES motifs within primary miRNAs (pri-miRNAs) and their relationship to the RIIIDa and RIIIDb domains of the DROSHA enzyme.