### Supplementary Tables

#### Table 1 - Basic statistics of the CLIP-seq datasets

Average number denotes the average number of motif-matched sites within the peak regions.

<table>
<thead>
<tr>
<th>RBP</th>
<th>motif</th>
<th>species[assembly]</th>
<th>experiment</th>
<th>number of motif site</th>
<th>average number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLD-1</td>
<td>AYUAAY</td>
<td>C.elegance(ce6)</td>
<td>PAR-CLIP</td>
<td>385</td>
<td>1.17</td>
</tr>
<tr>
<td>QKI</td>
<td>AYUAAY</td>
<td>H.sapiens(hg18)</td>
<td>PAR-CLIP</td>
<td>3054</td>
<td>1.26</td>
</tr>
<tr>
<td>Pum2</td>
<td>UGUANUA</td>
<td>H.sapiens(hg18)</td>
<td>PAR-CLIP</td>
<td>1327</td>
<td>1.054</td>
</tr>
<tr>
<td>SF2ASF</td>
<td>GAAGAA</td>
<td>H.sapiens(hg18)</td>
<td>HITS-CLIP</td>
<td>2721</td>
<td>1.2521</td>
</tr>
<tr>
<td>Nova</td>
<td>YCAY</td>
<td>M.musculus(mm9)</td>
<td>HITS-CLIP</td>
<td>24019</td>
<td>1.345</td>
</tr>
<tr>
<td>Lin28A</td>
<td>AAGNNG</td>
<td>M.musculus(mm9))</td>
<td>HITS-CLIP</td>
<td>28642</td>
<td>1.1164</td>
</tr>
<tr>
<td>FXR1</td>
<td>ACUK or WGGA</td>
<td>H.sapiens(hg18)</td>
<td>PAR-CLIP</td>
<td>2634</td>
<td>1.15</td>
</tr>
<tr>
<td>FXR2</td>
<td>ACUK or WGGA</td>
<td>H.sapiens(hg18)</td>
<td>PAR-CLIP</td>
<td>12886</td>
<td>1.2112</td>
</tr>
<tr>
<td>FMR1_7</td>
<td>ACUK or WGGA</td>
<td>H.sapiens(hg18)</td>
<td>PAR-CLIP</td>
<td>46826</td>
<td>1.43478</td>
</tr>
<tr>
<td>FMR1_1</td>
<td>ACUK or WGGA</td>
<td>H.sapiens(hg18)</td>
<td>PAR-CLIP</td>
<td>93678</td>
<td>1.616</td>
</tr>
</tbody>
</table>

#### Table 2 - The numbers of two known sequential motifs for the CLIP-seq data set of the FMRP family

<table>
<thead>
<tr>
<th>RBP</th>
<th>ACUK</th>
<th>WGGA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FXR1</td>
<td>2435</td>
<td>199</td>
<td>2634</td>
</tr>
<tr>
<td>FXR2</td>
<td>9829</td>
<td>3057</td>
<td>12886</td>
</tr>
<tr>
<td>FMR1_7</td>
<td>19159</td>
<td>27667</td>
<td>46826</td>
</tr>
<tr>
<td>FMR1_1</td>
<td>46364</td>
<td>47314</td>
<td>93678</td>
</tr>
</tbody>
</table>
Supplementary Figures

Supplementary Fig. 1  Dependence of the exterior loop, multibranch loop, and unstructured context on the maximal span $W$. The x-axis represents the maximal span $W$. The y-axis represents the averaged $p(i, \delta)$ over all the nucleotides.

A  

GLD-1

B  

QKI

C  

PUM2

D  

SRSF1
Supplementary Fig. 2  Dependence of the distribution of P-score on the maximal span $W$. The distribution of P-scores with various maximal span $W$ for (A) GLD-1 (unstructured), (B) QKI (unstructured), (C) Pum2 (hairpin), (D) SRSF1 (unstructured), (E) Nova (unstructured), (F) Lin28A (internal), (G) FXR1 (ACUK) (unstructured), (H) FXR1 (WGGA) (stem), (I) FXR2 (ACUK) (unstructured), (J) FXR2 (WGGA) (stem), (K) FMR1 (ACUK) (stem), (L) FMR1 (WGGA) (stem), (M) FMR1 (ACUK) (stem), and (N) FMR1 (WGGA) (hairpin). The x-axis represents nucleotide positions. The y-axis is P-score. The black box represents sequential motif sites. For each RBP, only the distribution for structural context with highest P-score at $W = 200$ are shown.
Supplementary Fig. 3 The distribution of the P-scores for each RBP. The x-axis represents nucleotide positions, and the y-axis represents P-score of ±20 bases around the sequential motif site. The position 0 denotes the start position of the sequential motif. Positive P-scores for each structural context indicate that the positions tend to prefer the structural context. The black box represents the sequential motif site. The dotted lines are the corrected significance level of Bonferroni correction (α = 0.05). Each panel represents the distribution of P-score for (A) GLD-1, (B) QKI, (C) Pum2, (D) SRSF1, (E) Nova, (F) Lin28A, (G) FXR1(ACUK), (H) FXR1(WGGA), (I) FXR2(ACUK), (J) FXR2(WGGA), (K) FMR1_7(ACUK), (L) FMR1_7(WGGA), (M) FMR1_1(ACUK), and (N) FMR1_1(WGGA).
Supplementary Fig. 4 The distribution of the P-scores for the unbound datasets. The x-axis represents nucleotide positions, and the y-axis represents P-score of ±20 bases around the sequential motif site. The position 0 denotes the start position of the sequential motif. Positive P-scores for each structural context indicate that the positions tend to prefer the structural context. The black box represents the sequential motif site. The dotted lines are the corrected significance level of Bonferroni correction ($\alpha = 0.05$). Each panel represents the distribution of P-score for (A) GLD-1, (B) QKI, (C) Pum2, (D) SRSF1, (E) Nova, (F) Lin28A, (G) FXR1(ACUK), (H) FXR1(WGGA), (I) FXR2(ACUK), (J) FXR2(WGGA), (K) FMR1_7(ACUK), (L) FMR1_7(WGGA), (M) FMR1_1(ACUK), and (N) FMR1_1(WGGA).
Supplementary Fig. 5  The distribution of the P-scores for the shuffled datasets. The x-axis represents nucleotide positions, and the y-axis represents P-score of ±20 bases around the sequential motif site. The position 0 denotes the start position of the sequential motif. Positive P-scores for each structural context indicate that the positions tend to prefer the structural context. The black box represents the sequential motif site. The dotted lines are the corrected significance level of Bonferroni correction ($\alpha = 0.05$). Each panel represents the distribution of P-score for (A) GLD-1, (B) QKI, (C) Pum2, (D) SRSF1, (E) Nova, (F) Lin28A, (G) FXR1(ACUK), (H) FXR1(WGGA), (I) FXR2(ACUK), (J) FXR2(WGGA), (K) FMR1_7(ACUK), (L) FMR1_7(WGGA), (M) FMR1_1(ACUK), and (N) FMR1_1(WGGA).
Supplementary Fig. 6  The nucleotide composition around the each RBP-bound sites. The nucleotide compositions of ±20 bases around the RBP-bound sites for (A) GLD-1, (B) QKI, (C) Pum2, (D) SRSF1, (E) Nova, (F) Lin28A, (G) FXR1(ACUK), (H) FXR1(WGGA), (I) FXR2(ACUK), (J) FXR2(WGGA), (K) FMR1_7(ACUK), (L) FMR1_7(WGGA), (M) FMR1_1(ACUK), and (N) FMR1_1(WGGA). The x-axis represents the nucleotide position, and the y-axis is the proportion of each nucleotide. The black box represents the sequential motif site.
Supplementary Fig.7 The comparison of P-scores of the positive datasets with P-scores of the shuffled and partially shuffled datasets. The shuffled, the partially shuffled (±5), and the partially shuffled (±10) datasets are represented by 0, 5, and 10, respectively. The x-axis represents the nucleotide position, and the y-axis represents P-score of (A) GLD-1 (unstructured), (B) QKI (unstructured), (C) Pum2 (hairpin), (D) SRSF1 (unstructured), (E) Nova (unstructured), (F) Lin28A (internal), (G) FXR1 (ACUK) (unstructured), (H) FXR1 (WGGA) (stem), (I) FXR2 (ACUK) (unstructured), (J) FXR2 (WGGA) (stem), (K) FMR1,7 (ACUK) (stem), (L) FMR1,7 (WGGA) (stem), (M) FMR1,1 (ACUK) (stem), and (N) FMR1,1 (WGGA) (hairpin). The black box is the RBP-bound sites, and the horizontal dotted line the corrected significance level of Bonferroni correction. The vertical dotted lines indicate the ±5 or 10 nt of RBP-bound sites. Note that, only the distribution for structural context with highest P-score are shown for each RBP.

Supplementary Fig.8 The dependence of structural profiles on the truncated length. The x-axis represents the truncated length. The y-axis represents the Pearson correlation coefficient between the structural profiles of the original sequence and those of the truncated sequences.
Supplementary Fig.9  The W-sensitivities of exterior loop, multibranch loop, and unstructured contexts for CLIP-seq datasets. The y-axis represents the W-sensitivity. The low W-sensitivity means that the highest P-score at $W = 30$ is larger than that at $W = 400$, and vice versa. When W-sensitivity ($\delta$) equals zero, the structural context $\delta$ is completely insensitive to the maximal span.