Comprehensive analysis of RNA-protein interactions by high-throughput sequencing–RNA affinity profiling

Jacob M Tome1,3, Abdullah Ozer1,3, John M Pagano1, Dan Gheba2, Gary P Schroth2 & John T Lis1

RNA-protein interactions play critical roles in gene regulation, but methods to quantitatively analyze these interactions at a large scale are lacking. We have developed a high-throughput sequencing–RNA affinity profiling (HiTS-RAP) assay by adapting a high-throughput DNA sequencer to quantify the binding of fluorescently labeled protein to millions of RNAs anchored to sequenced cDNA templates. Using HiTS-RAP, we measured the affinity of mutated libraries of GFP-binding and NELF-E–binding aptamers to their respective targets and identified critical regions of interaction. Mutations additively affected the affinity of the NELF-E–binding aptamer, whose interaction depended mainly on a single-stranded RNA motif, but not that of the GFP aptamer, whose interaction depended primarily on secondary structure.

RNA-protein interactions are ubiquitous in biology and critical at many regulatory steps of gene expression and stages of organismal development1. These interactions vary in their mechanism of action; for example, the XIST long noncoding RNA (lncRNA) guides regulatory proteins to chromatin, the HOTAIR lncRNA acts as a scaffold for the assembly of regulatory complexes2, and nascent RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3.RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3. RNAs can regulate their own transcription by coordinating the transition of RNA polymerase (RNAP) II into productive elongation3.

The Illumina platform allows simultaneous sequencing of hundreds of millions of DNA clusters, each derived from the amplification of a single molecule with primers that are covalently linked to a glass flow cell18. These sequencers can also be used to image and analyze the binding of fluorescently labeled proteins to DNA clusters, as demonstrated with the high-throughput sequencing–fluorescent ligand interaction profiling (HiTS-FLIP) protocol19. A corresponding assay for RNA binding to protein has not been realized owing to the difficulty of converting DNA clusters into RNA that is retained at the cluster20.

We have developed the HiTS-RAP assay, which transcribes DNA at each cluster on an Illumina flow cell, and used it to measure RNA aptamer affinities at very large scale. We measured dissociation constants for 1,875 mutants of the GFP aptamer (GFPapt)21 and 9,832 mutants of NELFapt22, an aptamer that binds Drosophila NELF-E (negative elongation factor E). NELF-E is an RNA-binding subunit of a protein complex involved in maintaining RNAP II pausing23–26. Analysis of these data allowed us to determine sequence and structural features that are most important for mediating protein interactions with these two aptamers. These results demotrate that HiTS-RAP can be used to characterize features of RNAs that determine their interactions with proteins. More generally, this work establishes HiTS-RAP as an effective and large-scale means of quantitatively evaluating interactions between RNAs and other molecules. We anticipate that it will be used to characterize biologically important RNA-binding proteins and to identify and optimize RNAs for a number of applications in biotechnology and medicine.

RESULTS
Tus stably halts transcription
To stably retain RNA transcribed from DNA on the Illumina flow cell at each DNA cluster, we used the Escherichia coli replication terminator protein Tus as a roadblock to transcription. Tus halts DNA replication at sites of replication termination in an orientation-specific manner by binding to the 32-bp Ter sequence element with high affinity, specificity and stability27–30. In addition, Tus binding stops RNAPs from transcribing through Ter sites in the nonpermissive orientation, resulting in either terminated or halted transcription27,31. We found that truncated transcripts,

1Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA. 2Illumina, Inc., San Diego, California, USA. 3These authors contributed equally to this work. Correspondence should be addressed to J.T.L. (johnlis@cornell.edu).

RECEIVED 12 FEBRUARY; ACCEPTED 28 APRIL; PUBLISHED ONLINE 8 MAY 2014; DOI:10.1038/NMETH.2970
many of which remained anchored to the DNA, were produced by T7 RNAP only when Tus was bound to DNA in the nonpermissive orientation (Supplementary Fig. 1a). An electrophoretic mobility shift assay (EMSA) with radiolabeled DNA showed that after transcription halting, nearly every DNA was engaged in a complex of intermediate mobility containing an RNA transcript (Fig. 1a and Supplementary Fig. 1b). Thus, a nonpermissive Tus-Ter complex halted T7 RNAP transcription upstream of the Ter site, resulting in an RNA transcript that was bound to the DNA by the polymerase (Supplementary Note 1).

We used an RNA aptamer that has high affinity and specificity for GFP and EGFP to develop HiTS-RAP. As an initial test, we attached aptamer template DNAs to beads and generated halted transcription complexes. EGFP bound beads with halted GFPapt RNA sequence (Fig. 1b). A bead experiment using single-round transcription showed that this interaction was due to the full-length GFPapt RNA (Supplementary Fig. 2).

**RNA tethering on an Illumina sequencer**

To couple Tus-dependent halting of T7 RNAP with sequencing on an Illumina GAIIx instrument, we constructed DNA libraries containing target aptamer template flanked upstream by a T7 promoter and downstream by the Ter sequence (Supplementary Fig. 3). All steps of HiTS-RAP—DNA sequencing, new second DNA-strand generation, Tus binding, transcription and halting, and probing with fluorescently labeled protein—are carried out automatically (Fig. 2a and Supplementary Software). Illumina software is used to measure the fluorescence intensity of bound mOrange fusion protein at each cluster. The total-internal-reflection
fluorescence microscopy used by the sequencer enables equilibrium measurements by imaging only at the glass surface where protein binds clusters, effectively ignoring excess protein in solution.18,19

We assayed a population of GFPapt point mutants and control RNA for their affinity to EGFP-mOrange using HiTS-RAP. EGFP-mOrange was used because EGFP is not detectable with the optics of the sequencer.18,33 The vast majority of GFPapt DNA clusters produced halted RNA capable of binding to EGFP-mOrange, whereas a lane containing clusters encoding the negative-control SRB-2 aptamer showed no appreciable binding in HiTS-RAP (Fig. 2b). Extrapolating from repeated measurements of GFPapt cluster fluorescence intensity at high protein concentration, we estimate that the assay can measure signal above background after transcription for 48 cycles, or 72 h, given an approximate cycle time of 1.5 h (Online Methods and Supplementary Fig. 4). Thus, the halted transcription complexes were sufficiently stable to carry out the several sequential measurements necessary to determine dissociation constants (Kd values).

### Measuring Kd values for GFP aptamer variants

We performed HiTS-RAP on a flow cell with three lanes containing the GFPapt template and a population of point mutants generated by many rounds of PCR. To measure the Kd values of interactions with EGFP-mOrange, we used seven protein concentrations in fivefold increments from 0.04 to 625 nM. We determined affinities for 1,875 mutant GFPapt sequences (Supplementary Table 1) from three lanes on the flow cell (Online Methods). We measured a Kd of 4.27 ± 1.11 nM (geometric mean multiplied/divided by geometric s.d.; ref. 34) for the EGFP-GFPapt interaction, a value in agreement with its published affinity of 5–15 nM Kd (ref. 21), whereas the SRB-2 aptamer negative control showed no appreciable binding in HiTS-RAP (Fig. 2c).

Most GFPapt mutants did not differ substantially in sequence from the canonical aptamer and bound EGFP with similar affinity. However, some showed altered affinity. To verify measured Kd values, we picked eight such mutants and measured their binding affinity by EMSA (Fig. 2d) and Supplementary Fig. 5a. HiTS-RAP and EMSA-measured affinities were well correlated (Pearson r2 = 0.64; Supplementary Fig. 5b and Supplementary Table 2). Some of the GFPapt mutants that we identified were represented by only a small fraction of the total clusters on the flow cell, demonstrating that HiTS-RAP could measure Kd values of even low–copy number sequences in a library (Supplementary Table 1).

### Comprehensive structure-function analysis of GFPapt

We used our library of mutants to carry out an in-depth analysis of the 82 nt of the 84-nt GFPapt. Of the 246 possible single-base substitutions, 236 were present in our data set (Table 3).

**Figure 3** Analysis of a GFPapt library by HiTS-RAP. (a) GFP-binding affinities of all 236 single-point mutants of GFPapt. Mutations at each position are color coded by nucleotide. Wild-type GFPapt binding affinity is indicated by the colored dashed line, and the sequence is shown at the bottom of the graph. Error bars, s.d. in log(Kd); see also Supplementary Table 1. 175 mutants qualify as binding in all three lanes. Single-point mutants that qualify as not binding (n.b.), and are thus assigned an affinity of 125 nM in at least one lane are plotted with open circles with no error bars. Those that do not bind in all lanes are at the top of the plot. (b) Predicted secondary structure of GFPapt. GFPapt is predicted to fold into a three–stem-loop structure connected by a central three-way junction. Each position is colored by the average absolute effect (log2(Kd_mut/Kd_wt)) of all its measured mutants. Mutations that qualify as not binding are assigned an affinity of 125 nM for this position. Where the average effects are greater than 4 (>16-fold effect in affinity) are colored red. Most mutations have a negative effect or less than a twofold-positive effect on binding affinity, except CS8U and U60A (asterisks). (c) Correlation between measured and predicted effects of GFPapt double mutants. Measured log10(Kd_mut1/Kd_wt) is plotted against the value predicted on the basis of single-point mutants (log10(Kd_mut1/Kd_wt) + log10(Kd_mut2/Kd_wt)). Points are colored according to the difference between measured and predicted effects. There is a positive but small correlation (r2 = 0.10) between the measured and predicted effects, and they differ as much as 1.98 (~100-fold).

#### Supplementary Material

- **Figure 1**: Expected secondary structure of GFPapt.
- **Figure 2**: Measurement of GFPapt binding. (a) Exponential growth of fluorescent signal from transient expression in EBV-LCL. The signal reaches a maximum at about 48 hpt. (b) Fluorescence microscopy used by the sequencer enables determination of the early stage of transcription (4 or 8 hpt). (c) Fluorescence microscopy used by the sequencer enables determination of the late stage of transcription (24 or 48 hpt). (d) Correlation between measured and predicted values of 100,000 random vectors. (e) Fluorescence microscopy used by the sequencer enables determination of the late stage of transcription (24 or 48 hpt). (f) Fluorescence microscopy used by the sequencer enables determination of the late stage of transcription (24 or 48 hpt).

#### Table 1

| Position | Effect | Nucleotide |
|----------|--------|------------|
| 1        | +      | A          |
| 2        | –      | C          |
| 3        | +      | G          |
| 4        | –      | C          |
| 5        | +      | A          |

#### Table 2

| Position | Effect | Nucleotide |
|----------|--------|------------|
| 1        | +      | A          |
| 2        | –      | C          |
| 3        | +      | G          |
| 4        | –      | C          |
| 5        | +      | A          |

#### Table 3

| Position | Effect | Nucleotide |
|----------|--------|------------|
| 1        | +      | A          |
| 2        | –      | C          |
| 3        | +      | G          |
| 4        | –      | C          |
| 5        | +      | A          |

#### Notes

1. Supplementary methods and results are provided in the Supplementary Material.
2. The data set used in the analysis is available at the following URL.
3. The authors declare no competing financial interests.
4. Correspondence and requests for materials should be addressed to S.J.B. (s.j.baker@uw.edu).
5. All animals were handled in accordance with institutional guidelines and approved by the institutional animal care and use committee.
6. The authors thank J. F. Coggeshall, B. A. Smerdon, and M. T. Kozlov for critical reading of the manuscript.
7. The research was supported by the National Institutes of Health and the National Science Foundation.
8. Supplementary Information is available on the Nature Methods website.
9. Supplementary video is available on the Nature Methods website.
10. Supplementary data are available on the Nature Methods website.
11. Supplementary methods and results are provided in the Supplementary Material.
12. The authors thank J. F. Coggeshall, B. A. Smerdon, and M. T. Kozlov for critical reading of the manuscript.
13. The research was supported by the National Institutes of Health and the National Science Foundation.
14. Supplementary Information is available on the Nature Methods website.
15. Supplementary video is available on the Nature Methods website.
16. Supplementary data are available on the Nature Methods website.
17. The authors thank J. F. Coggeshall, B. A. Smerdon, and M. T. Kozlov for critical reading of the manuscript.
18. The research was supported by the National Institutes of Health and the National Science Foundation.
19. Supplementary Information is available on the Nature Methods website.
20. Supplementary video is available on the Nature Methods website.
21. Supplementary data are available on the Nature Methods website.
22. The authors thank J. F. Coggeshall, B. A. Smerdon, and M. T. Kozlov for critical reading of the manuscript.
23. The research was supported by the National Institutes of Health and the National Science Foundation.
24. Supplementary Information is available on the Nature Methods website.
25. Supplementary video is available on the Nature Methods website.
26. Supplementary data are available on the Nature Methods website.
27. The authors thank J. F. Coggeshall, B. A. Smerdon, and M. T. Kozlov for critical reading of the manuscript.
28. The research was supported by the National Institutes of Health and the National Science Foundation.
Most mutations show less-than-twofold effects (\(\log_2(K_d,\text{mut}/K_d,\text{wt}) \leq 1\)) on binding affinity. The minimal aptamer (NELFapt min) is indicated with a dashed box. (c) Correlation between measured and predicted effects of NELFapt double mutants (\(n = 2,442\)). The measured fold effect of double mutants is plotted against the fold effect predicted by single mutations as in Figure 3c. In this case, there is a strong positive correlation (\(r^2 = 0.87\)) between measured and predicted fold effect.

The \(K_d\) values of both of these mutants by EMSA (Supplementary Fig. 5). In contrast, two single-point mutants with \(K_d > 125\) nM, G68C and G46C, caused gross alterations in the predicted structure of GFPapt consistent with their substantial adverse effects on binding affinity (Supplementary Fig. 6b).

In agreement with the original study reporting GFPapt\(^2\), the majority of nucleotide positions with a high impact on binding were located in either stem-loop 2 or 3 (Fig. 3b).

We also examined the relationship between the affinity of double mutants and their corresponding single mutants. We identified 181 cases in which the double mutant and both corresponding single mutants had high-confidence EGFP-binding affinities (Supplementary Table 4). If two single mutations affect binding independently, the effect of the double mutant on binding should be additive. However, cooperative or compensatory effects on binding should cause the measured effect of the double mutant to deviate from the additive expectation. To test for this, we used a metric analogous to \(\Delta G\) (difference in Gibbs free-energy change) to represent the effect of a mutation: \(\log(K_d,\text{mut}/K_d,\text{wt})\), the log ratio of \(K_d\) values for the mutant and canonical aptamer. There was only a weak correlation (\(r^2 = 0.10\)) between the measured effects of the GFPapt double mutants and the prediction under an additive model (Fig. 3c).

Double mutants with substantially higher or lower affinity than predicted can highlight features important for interaction with EGFP (Supplementary Table 4). For example, A23G-U34C reconstituted an AU base pair as a GC base pair in the predicted secondary structure and bound with higher affinity than predicted, whereas the same U34C mutation in combination with C58U failed to rectify the structural perturbation caused by U34C and had lower affinity than predicted (Supplementary Fig. 7). Overall, mutational analysis and structural predictions indicated that the interaction between the GFP aptamer and its target is complex in nature, depending upon an intricate structure dictated by its sequence.

High-throughput affinity profiling of Drosophila NELF-E

We used HiTS-RAP to examine the interaction between the RNA aptamer NELFapt and NELF-E, an RNA-binding protein with a highly conserved RNA-recognition motif\(^2\). NELF-E recognizes the NELF-E–binding element (NBE), a 7-nt motif (CUGAGGA) located within a putative kink-turn (k-turn) motif in NELFapt that forms a sharp bend between two stems to present the NBE in a single-stranded loop region\(^2\). To characterize this interaction, we mutated NELFapt through error-prone PCR and performed HiTS-RAP to probe NELF-E binding using a single lane of an Illumina GAIIx flow cell. We measured high-confidence binding affinities for 9,832 mutants (Supplementary Table 5). The \(K_d\) for full-length NELFapt was measured to be 5.2 nM, which was in good agreement with EMSA measurements (Supplementary Fig. 8).

All 210 possible single-base substitutions within the 70-nucleotide NELFapt were identified (Supplementary Table 6). 206 with high-confidence fits for \(K_d\) values (Fig. 4a). The majority of single-point mutations did not have a notable effect on binding affinity, and none had substantially higher affinity than the canonical aptamer. Mutations within the NBE were among the most disruptive to the interaction between the aptamer and NELF-E (Fig. 4b), with measured affinities as low as 76 nM for A43C. Interestingly, bases in the loop region outside the NBE located at positions G58 and A59. The deleterious effects of these two mutations were confirmed by EMSA (Supplementary Fig. 8). These residues likely interact with G45 and A46 through the non-Watson-Crick base-pairing characteristic of k-turn structures\(^3\). The observed effects of the single-point mutations strongly support a k-turn structure of this aptamer with the presentation of the NBE as a single-stranded loop.

**Figure 4** Analysis of NELFapt by HiTS-RAP. (a) NELF-E–binding affinities of 206 single-point mutants of NELFapt. Mutations and the canonical aptamer sequence shown at the bottom of the graph are colored as in Figure 3a. (b) Predicted secondary structure of NELFapt, showing two stem-loops connected via a loop-stem-loop structure. The structure is drawn and colored as in Figure 3b. Most mutations show less-than-twofold effects (\(\log_2(K_d,\text{mut}/K_d,\text{wt}) \leq 1\)) on binding affinity. The minimal aptamer (NELFapt min) is indicated with a dashed box. (c) Correlation between measured and predicted effects of NELFapt double mutants (\(n = 2,442\)). The measured fold effect of double mutants is plotted against the fold effect predicted by single mutations as in Figure 3c. In this case, there is a strong positive correlation (\(r^2 = 0.87\)) between measured and predicted fold effect.
To assess the interplay of different nucleotides, we compared affinities of mutant NELF-E aptamers with single- and double-point mutations. We identified 2,442 double mutants with high-confidence $K_d$ values (Supplementary Table 7). In contrast to the results with GFPapt, the affinities of NELFapt double mutants were predicted well by the affinities of the individual single mutants ($r^2 = 0.87$; Fig. 4c). Most double mutants had small effects on affinity and were well predicted by an additive model. Many that did not follow this trend are notable. Some were compensatory: for example, both A39G-U61C and A39U-U61A bound NELF-E better than predicted by their corresponding single mutants, presumably because the double mutants reconstituted a predicted AU base pair as GC and UA base pairs, respectively, on a critical stem near the NBE (Supplementary Fig. 9a). Others, such as A39G-G63A, bound with lower affinity than predicted by individual mutations because they may have resulted in sequestration of the NBE within a double-stranded region (Supplementary Fig. 9b), which was shown to be deleterious for NELF-E interaction$^2$.2. Our analysis showed that most mutations within NELFapt affected its interaction with NELF-E in an additive way, with only a few notable and informative exceptions.

**DISCUSSION**

HiTS-RAP provides a quantitative, high-throughput assay for RNA-protein interactions, which was lacking despite a need in the field. In fact, while this work was in review, Buenrostro et al.$^{36}$ published a similar technique that uses the halting of $E. coli$ RNAP by a biotin-streptavidin roadblock after single-round transcription to perform an assay very similar to ours.

After a normal sequencing run, all additional manipulations for HiTS-RAP are carried out automatically as part of the .xml recipe used by the sequencer, adding very little to cost and hands-on time. Because sequencing chemistry is identical across Illumina sequencers, HiTS-RAP is also compatible with the HiSeq and MiSeq instruments for increased data output or faster small-scale readout, respectively. Using HiTS-RAP, we have accurately measured the affinities of libraries of aptamers that bind EGFP and NELF-E, providing structural insight and identifying regions critical for RNA-protein interaction. Careful filtering minimized the contribution of sequencing errors (Online Methods and Supplementary Note 2), as suggested by the close correspondence of $K_d$ values measured by HiTS-RAP and EMSA.

In this work, we used HiTS-RAP to gain insight into the binding of two RNA aptamers by quantitatively measuring the affinity of a large number of mutants. This allowed us to identify critical features of both of these very different RNA-protein interactions. We found that most bases within GFPapt were essential for its interaction with EGFP, and when multiple mutations were present, these often confounded each other’s effects, consistent with GFPapt interacting with EGFP as an intricate three-dimensional structure. However, we did identify two single-point mutations within the aptamer that conferred several-fold higher affinity, demonstrating the ability of HiTS-RAP to optimize already extensively selected aptamers (Supplementary Note 3). Our mutagenesis of NELFapt showed that only a few features of the aptamer were indispensable for binding, illustrating the ability of HiTS-RAP to identify minimal aptamers. In addition, we showed that sequence-specific recognition of the single-stranded NBE by the NELF-E RNA-recognition motif was necessary but not sufficient for this interaction: the structural context of NBE presentation on a single-stranded loop within a k-turn was also important. This finding has implications in understanding how this critical transcriptional regulatory protein interacts with RNA.

The HiTS-RAP principle of halting transcription with Tus binding could be used to convert platforms and assays currently restricted to DNA, such as microarrays, to measure the properties of RNA as well. Additionally, although we have used HiTS-RAP to measure affinities of RNAs to a protein, its utility extends to any entity that can be fluorescently labeled (for example, small molecules and peptides). Last, DNA libraries tailored for many applications, including random sequence libraries, aptamer libraries generated by SELEX, random genomic fragments or targeted genomic libraries (such as nascent RNA, mRNA, noncoding RNA, enhancer regions or pre-mRNA) can be easily adapted or constructed de novo for HiTS-RAP. Thus, HiTS-RAP could facilitate genome-wide, direct, quantitative measurement of RNA affinity for regulatory proteins.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** Addgene: 53302 (EGFP-mOrange pHis-//), 53303 (NELFE-mOrange pHis-//), 53304 (mOrange pHis-//) and 53305 (Tus pGST-//). NCBI Sequence Read Archive: SRP041429 (raw HiTS-RAP data).

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank R. Friedman (Institut Pasteur) for providing scripts to extract intensity data, B. Moharnty (Medical University of South Carolina) for providing vectors containing the tus gene, K. Szeto and D. Shalloway for advice on data analysis, W. Zipfel and A. Singh for help in understanding the optics of the GAIDa, C.T. Waters and H. Kwak for bioinformatics advice, the Cornell sequencing core facility for help in learning the GAIx, and H. Craighead (Cornell University) and the members of the Li lab for helpful discussions on experimental design and the manuscript. This work was supported by US National Institutes of Health grants GM090320 and DA030329 to J.T.L.

**AUTHOR CONTRIBUTIONS**

Initial idea: A.O. and J.T.L. Experimental design: J.M.T., A.O., D.G., G.P.S. and J.T.L. Experimental implementation: J.M.T., A.O. and J.M.P. HiTS-RAP experiments and bioinformatics analysis: J.M.T. Project coordination: J.T.L. Paper writing: J.M.T., A.O., J.M.P. and J.T.L.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Lee, J.T. Epigenetic regulation by long noncoding RNAs. Science 338, 1435–1439 (2012).
2. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166 (2012).
3. Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell 153, 855–868 (2013).
4. Keefe, A.D., Pai, S. & Ellington, A. Aptamers as therapeutics. Nat. Rev. Drug Discov. 9, 537–550 (2010).
5. Germer, K., Leonard, M. & Zhang, X. RNA aptamers and their therapeutic and diagnostic applications. Int. J. Biochem. Mol. Biol. 4, 27–40 (2013).
ARTICLES

6. Shi, H., Hoffman, B.E. & Lis, J.T. RNA aptamers as effective protein antagonists in a multicellular organism. *Proc. Natl. Acad. Sci. USA* **96**, 10033–10038 (1999).

7. König, J., Zarnack, K., Luscombe, N.M. & Ule, J. Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.* **13**, 77–83 (2011).

8. Wong, I. & Lohman, T.M. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc. Natl. Acad. Sci. USA* **90**, 5428–5432 (1993).

9. Ryder, S.P., Recht, M.I. & Williamson, J.R. Quantitative analysis of RNA-protein interactions by gel mobility shift. *Methods Mol. Biol.* **488**, 99–115 (2008).

10. Pagano, J.M., Clingman, C.C. & Ryder, S.P. Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. *RNA* **17**, 14–20 (2011).

11. Salim, N.N. & Feig, A.L. Isothermal titration calorimetry of RNA. *Methods* **47**, 198–205 (2009).

12. Katsamba, P.S., Park, S. & Laird-Offringa, I.A. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* **26**, 95–104 (2002).

13. Campbell, Z.T. et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep.* **1**, 570–581 (2012).

14. Martin, L. et al. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat. Methods* **9**, 1192–1194 (2012).

15. Tenenbaum, S.A., Carson, C.C., Lager, P.J. & Keene, J.D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl. Acad. Sci. USA* **97**, 14085–14090 (2000).

16. Licatalosi, D.D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).

17. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

18. Bentley, D.R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–58 (2008).

19. Nutiu, R. et al. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat. Biotechnol.* **29**, 659–664 (2011).

20. Evanko, D. Next-generation protein binding. *Nat. Methods* **8**, 619 (2011).

21. Shui, B. et al. RNA aptamers that functionally interact with green fluorescent protein and its derivatives. *Nucleic Acids Res.* **40**, e39 (2012).

22. Pagano, J.M. et al. Defining NELF-E RNA binding in HIV-1 and promoter-proximal pause regions. *PloS Genet.* **10**, e1004090 (2014).

23. Wu, C.-H. et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17**, 1402–1414 (2003).

24. Wu, C.-H. et al. Molecular characterization of *Drosophila* NELF. *Nucleic Acids Res.* **33**, 1269–1279 (2005).

25. Yamaguchi, Y. et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).

26. Yamaguchi, Y., Inukai, N., Narita, T., Wada, T. & Handa, H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Mol. Cell. Biol.* **22**, 2918–2927 (2002).

27. Mohanty, B.K., Sahoo, T. & Bastia, D. The relationship between sequence-specific termination of DNA replication and transcription. *EMBO J.* **15**, 2530–2539 (1996).

28. Kamada, K., Horiiuchi, T., Ohsumi, K., Shimamoto, N. & Morikawa, K. Structure of a replication-terminator protein complexed with DNA. *Nature* **383**, 598–603 (1996).

29. Mulugu, S. et al. Mechanism of termination of DNA replication of *Escherichia coli* involves helicase-contrahelicase interaction. *Proc. Natl. Acad. Sci. USA* **98**, 9569–9574 (2001).

30. Mulcair, M.D. et al. A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*. *Cell* **125**, 1309–1319 (2006).

31. Guajardo, R. & Sousa, R. Characterization of the effects of *Escherichia coli* replication terminator protein (Tus) on transcription reveals dynamic nature of the Tus block to transcription complex progression. *Nucleic Acids Res.* **27**, 2814–2824 (1999).

32. Holeman, L.A., Robinson, S.L., Szostak, J.W. & Wilson, C. Isolation and characterization of fluorophore-binding RNA aptamers. *Folding Des.* **3**, 423–431 (1998).

33. Shaner, N.C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).

34. Limpert, E. & Stahel, W.A. Problems with using the normal distribution—and ways to improve quality and efficiency of data analysis. *PLoS ONE* **6**, e21403 (2011).

35. Klein, D.J., Schmeing, T.M., Moore, P.B. & Steitz, T.A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).

36. Buenrostro, J.D. et al. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* doi:10.1038/nbt.2880 (13 April 2014).
**ONLINE METHODS**

**Purification of proteins.** The gene encoding Tus protein was PCR amplified from a vector provided by B. Mohanty (Medical University of South Carolina) and then inserted between BamHI and XhoI restriction sites in the expression vector pGST-parallel. GST-Tus was then overexpressed in BL21 (DE3) RIL bacteria for 4 h after induction with 1 mM IPTG at 37 °C. The protein was purified using glutathione coupled agarose resin (Pierce), dialyzed into 40 mM Tris- HCl, pH 7.5, 40 mM NaCl, 2 mM EDTA, and 10 mM β-mercaptoethanol, allowed to mature at 4 °C for 1 month, mixed with an equal volume of 80% glycerol, and stored at −80 °C.

An EGFP-mOrange construct with a 4× GGGS flexible linker was produced by overlap-extension PCR and inserted between BamHI and XhoI restriction sites of the expression vector pHis-parallel. 6xHis-EGFP-mOrange was overexpressed in BL21 (DE3) RIL bacteria, as done for GST-Tus. The protein was purified using Ni-NTA-coupled agarose resin (Pierce), dialyzed into 10 mM Tris- HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, and 5 mM β-mercaptoethanol, allowed to mature at 4 °C for 1 month, mixed with an equal volume of 80% glycerol, and stored at −80 °C.

NELF-E-mOrange was made following a protocol similar to EGFP-mOrange. mOrange was first PCR amplified with BglII and BamHI sites at either end, digested with those enzymes, and inserted into the BamHI site of pHis-parallel, conserving its entire multiple cloning site. NELF-E was then subcloned from pHis-parallel into the mOrange containing pHis-parallel using EcoRI and SpeI restriction sites. Expression and purification were carried out exactly as for EGFP-mOrange.

**Library preparation.** Supplementary Figure 3 shows a schematic of a HiTS-RAP template. For GFpapt, mutations were introduced by PCR amplification of the template over a total of ~100 cycles. Realizing that PCR-induced mutations were limited and the throughput of the HiTS-RAP assay could accommodate a far greater number of mutants, we subjected NELFapt to a greater number of mutants, we subjected NELFapt to a greater number of mutations. DNA templates were prepared for sequencing and transcription halting by sequential addition of primers by PCR. First, a T7 promoter was added to the 5′ end of the GFP, sulforhodamine B (SRB), and NELF-E binding aptamers, and the Illumina sequencing primer site was added to the 3′ ends. Then adaptors complementary to oligos on the Illumina flow cell were added to either end along with a Ter site immediately 3′ of the Illumina sequencing primer site. The Ter site was positioned 3′ of the Illumina sequencing primer site to ensure that the target RNA is fully emerged from the polymerase upon halting and so that sequencing begins with the target RNA. A third PCR step was used to add adaptors for the 454 Life Sciences sequencing platform for templates used to make polystyrene beads with covalently linked halting templates. Sequences of primers used in this work are listed in Supplementary Table 8.

Final halting templates for HiTS-RAP (T7 RNAP promoter in *italics*, aptamer sequence underlined, *Ter* sequence in *bold*) are as follows.

The GFpapt halting template used for GFpapt was 5′-CAA GCAGAAGAGCCGATCGGCTGATATACGACGACTCA CTATAAGGGAATTGATCCATCTACGAAATTTCGCTTG GACTGCGGTGGGAGCACGAAACGTCGTGGCGCAATTGG CTATA GATAATACGACTCA ACTATA GGGAATGGATCCACATCTACGAATTCGGAAC...
beads were washed with GFPapt binding buffer (1× PBS, 5 mM MgCl₂, 0.01% Tween 20) and incubated with 1 µM 6xHis-EGFP. After 20 min of binding at room temperature, beads were washed with GFPapt binding buffer and imaged with a Zeiss Axioplan II epifluorescence microscope using a FITC/Fluo filter set (Chroma Technology cat. #41001). Both DIC and fluorescence images were taken.

Transcription halting on glutathione beads. DNA templates were prepared with the GFP aptamer template flanked on the 5’ end by a T7 promoter and an 11-nt C-less cassette and by two Ter sites on its 3’ end. First, these templates were bound to GST-Tus in 10× molar excess in 1× T7 transcription buffer for 30 min at 37 °C. The resulting Tus-DNA complexes were then incubated with glutathione-coupled agarose beads (Pierce) in 1× T7 transcription buffer for 30 min at 37 °C. An equal volume of 2× transcription reaction lacking CTP was then added to the resulting bead slurry, and transcription was allowed to proceed for 30 min at 37 °C. At this time, the one aliquot of beads was washed three times in 1× T7 transcription buffer and then mixed with an equal volume of 2× transcription reaction mix lacking polymerase but containing all four NTPs. Transcription was allowed to proceed for another 30 min at 37 °C. Both bead treatments were then washed with GFPapt binding buffer, incubated with 1 µM 6xHis-EGFP for 30 min at RT, washed again, and imaged as described before for 454 beads.

HiTS-RAP. A standard sequencing run was performed on an Illumina GAIIx using an 82-cycle read length for GFPapt and 126-cycle read length for NELFapt on a paired-end flow cell. The same .xml recipe used for the sequencing run included all subsequent steps to effect transcription halting and binding of mOrange-labeled protein to the sequenced DNA clusters, so that fresh solutions are added at once for all steps through transcription halting and then once for the binding curve. Thus, reagents for HiTS-RAP are loaded onto the Paired-End Module (PEM) twice. The recipe program used for the GFP run is included as Supplementary Software; it details the exact reagent delivery used for HiTS-RAP.

2.25 mL of each solution was loaded onto its own position on the PEM of the GAIIx. The .xml recipe delivered reagents in the proper sequence and set the flow cell temperature using the Peltier heater in the instrument. During each reaction or binding step, 75 µL of solution is flowed through each lane, and then the flow cell is incubated for 30 min to equilibrate. During the incubation, 15 µL of fresh solution is delivered to each lane every 5 min. The second strand generated during sequencing was stripped away, and 1 µM primer (IllumFORAdapt_T1_IllumFORSeq, Supplementary Table 8) for dsDNA regeneration annealed as per the standard Illumina protocol. Excess primer was then washed away with 1× NEB buffer 4 with 0.01% Tween 20 (New England Biolabs). DNA was then made double-stranded by flowing in a Klenow exo– enzyme reaction mix (1× NEB Buffer 4, 0.01% Tween 20, 0.2 mM dNTPs, Klenow exo– (New England Biolabs)) and incubating for 30 min at 37 °C. The flow cell containing dsDNA clusters was then equilibrated with 1× T7 transcription buffer. Tus was allowed to bind the DNA templates’ Ter elements by equilibrating with 1 µM GST-Tus in 1× T7 transcription buffer for 30 min at 37 °C. The flow cell was then equilibrated with a transcription reaction (1× T7 transcription buffer, 0.5 mM NTPs, T7 RNAP, YIP, SUPERase In, ~0.5 µM GST-Tus). Transcription and halting was allowed to proceed for 30 min at 37 °C.

After transcription halting, the flow cell was equilibrated with GFP or NELF (10 mM HEPES, pH 7.5, 100 mM NaCl, 25 mM KCl, 5 mM MgCl₂, 0.02% Tween 20) aptamer binding buffer at room temperature. It was imaged immediately, just as during sequencing, to measure the background intensity at every cluster. It was then equilibrated successively with increasing concentrations of mOrange-labeled protein in binding buffer and imaged. Imaging at each concentration was carried out in equilibrium binding after 30-min equilibration at room temperature. Concentrations of EGFP-mOrange varied from 0.04 nM to 625 nM, increasing in fivefold increments. The images in Figure 2b are from .tif files collected by the Illumina SCS during this binding. HiTS-RAP for NELF-E mOrange was identical to that for EGFP-mOrange, except that concentrations of protein varied from 0.064 to 1,000 nM.

Sequencing and data extraction. Sequencing, transcription, and protein binding are executed as a single run on a GAIIx. In the case of NELF-E, transcription and protein binding were carried out twice. The first binding curve was used for all analyses. Intensity data were collected and base-calling was done using the Illumina SCS version 2.9. During the run, .cif and .bcl files were saved, along with 5% of the raw .tif images taken. We used scripts generously provided by R. Friedman (Institut Pasteur) to extract intensities of mOrange fluorescence for every cluster from .cif file. This gives the coordinates of each cluster, together with its protein-binding intensities (in the T channel) and average intensity during sequencing. These data were then matched by cluster coordinates to their sequence from .qseq.txt files generated by the Illumina OLB version 1.9.4. Only clusters that passed filter were included. If a cluster contained a mutation (deviating from NELFapt or GFPapt), it was included only if the mutated base had a quality score greater than 25 (<0.005 probability of an incorrect base call) (Supplementary Software, GFPapt). A threshold of quality score >20 is common practice for SNP-calling algorithms.

Double-mutant analysis was restricted to base substitutions and to the region in the center of the aptamer (76 nt for GFPapt, 64 for NELFapt) so that the 3 bp at the ends could be used to ensure against insertions and deletions.

Loss of signal correction. After the GFP binding curve data were obtained, the flow cell was imaged nine successive times after re-equilibrating with 625 nM EGFP mOrange. After three of these cycles, imaging was not carried out, but rather the flow cell was allowed to sit in 625 nM EGFP-mOrange for the time that it would take for one equilibration and binding cycle. Thus, the nine imaging steps span the time that it would take for 12 equilibration and imaging steps. A similar rate of decay was observed between successive and staggered imaging cycles, indicating that most loss of signal is due to time rather than photobleaching caused by the number of times that the flow cell was imaged. We find that in the regime where our imaging takes place (i.e., where time is much less than the characteristic lifetime), the observed decay is described well by a linear approximation to exponential decay. Thus, we took the average intensities of clusters with canonical GFP aptamer sequence though the nine cycles imaged and used...
scipy.optimize.curve_fit to perform a weighted linear least-squares regression, using weights determined from the s.d. of the intensities at each time point. The equation used was

\[ I_{\text{observed}} = I_0\left(1 - \frac{t}{\tau}\right) \]

where \( I_{\text{observed}} \) is the measured mOrange intensity, \( I_0 \) is the initial intensity, \( t \) is the time in units of cycles, and \( \tau \) is the characteristic lifetime in cycles. \( I_0 \) and \( \tau \) were solved for the GFP aptamer clusters in each of the three lanes (each containing ~2.7 million clusters with canonical GFP aptamer sequence); the final characteristic lifetime is the average of these three fitted lifetimes. In this regime, halted complexes are decaying at a rate corresponding to a characteristic lifetime of 137 ± 7 cycles, or 206 ± 11 h (at an average cycle time of ~1.5 h). This lifetime simply describes the rate at which the halted transcription complexes are decaying, independent of the actual intensity. Given that the background intensity in these lanes is approximately 85, this means that fluorescence signal will be above background for 48 cycles, or 72 h.

The characteristic lifetime was used to apply a correction to each measured fluorescence intensity in binding curves, using the equation

\[ I_{\text{actual}} = \frac{I_{\text{observed}}}{1 - 0.00730 \times t} \]

where \( I_{\text{actual}} \) is the fluorescence intensity used in the fit for \( K_d \), \( I_{\text{observed}} \) is the measured intensity, 0.00730 is the decay rate in 1/cycles (the inverse of \( \tau \)), and \( t \) is the time since transcription, in cycles. Supplementary Figure 4a shows intensity data from three lanes, together with the same intensities after applying the correction factor.

**\( K_d \) calculation.** Sequential measurements of the corrected mOrange intensities at increasing concentrations give a binding curve for each cluster. Intensities were normalized for cluster size and position in the tile by dividing each intensity by the average sequencing intensity. This correction is applied so that all intensities from a lane can be averaged, to be representative of each sequence. Clusters with the same sequence in each lane were matched, and average binding curves were generated by taking the mean of their normalized intensities for each protein concentration (Supplementary Note 4). If more than ten clusters had a given sequence in a lane, this binding curve was fit to a Hill equation, solving the equilibrium dissociation constant (\( K_d \)) of the interaction. We have used the following Hill equation:

\[ I = b + \frac{m - b}{1 + \left(\frac{K_d}{C}\right)^n} \]

where \( b \) is the background fluorescence intensity at the cluster, \( m \) is the maximum fluorescence intensity, \( K_d \) is the dissociation constant, \( n \) is the Hill coefficient, and \( C \) is the concentration of target protein. The intensities measured by imaging at several different concentrations are then used to solve \( m, b, K_d \) and \( n \) in a weighted nonlinear least-squares regression. We have found that letting these four parameters vary gives reliable fits; this also means that the \( K_d \) that we solve is independent of the intensity values, so that it shows the inflection point in the binding curve.

Only fits for which the scipy.optimize.curve_fit algorithm of the NumPy Python package returned a variance of less than 1,000,000 were considered to be high confidence and used in these analyses.

The GFPapt run contained three lanes of GFPapt. To generate a single data set from all three lanes, we determined \( K_d \) values for every unique sequence represented by at least ten clusters in each lane by fitting to average binding curves. The \( K_d \) values for all unique sequences in the flow cell were then determined by geometrically averaging the fitted \( K_d \) values across the three lanes: therefore, the \( K_d \) values are reported as the average \( K_d \times \) (multiply or divide by) s.d.

The GFPapt data have the added complication of very low-background binding. Thus, there are sequences in this data set that do not bind GFP measurably, whereas for NELF-E, background binding for this RNA-binding protein is high enough that every sequence is expected to bind to some extent. To mitigate this problem, we considered as binding only sequences that show a 3% increase (determined in Supplementary Fig. 4b–d) between the first and last two measured intensities: all others were called as not binding, meaning that they effectively have a \( K_d \) greater than 125 nM. We set the limit at 125 nM because this was the second-highest concentration probed, so we do not expect to be able to measure affinities greater than this. For analysis of single mutants, any mutant scored as not binding on the basis of its intensity increase was assigned an affinity of 125 nM; this value was averaged with other real measurements from other lanes if it was called as binding there. This results in a \( K_d \) measurement of greater than that average. Sequences for which any lane was called as not binding are indicated in all tables. Such mutants were excluded from analysis of double mutants.

**EMSA of GFPapt and NELFapt and mutants.** EMSA was used to verify the HiTS-RAP–measured binding affinities of wild-type and mutant GFP-binding aptamers to EGFP. To this end, in vitro–transcribed aptamers were 3' end labeled with Alexa Fluor 647 hydrazide (Invitrogen) as described elsewhere and quantified by Qubit Fluorometer (Invitrogen). Fluorescently labeled aptamers were mixed with recombinant GST-EGFP protein at 25 °C for 45 min. The GST-EGFP concentration in the binding reaction was varied as a 2/3 dilution series starting from 500 nM, and a no-protein control was used. The final 20-µl binding reaction was composed of 1× PBS, 5 mM MgCl2, 0.4 U of SUPERase In, 1 µg of yeast tRNA, 0.005% NP-40, and 5 nM fluorescently labeled aptamer. After addition of bromocresol green containing 30% glycerol to 6% final glycerol concentration, binding reactions were loaded on a 6% polyacrylamide gel (0.5× TBE, 5 mM MgCl2) that was pre-equilibrated to 4 °C and prerun at 120 V for 10 min at 4 °C. Loaded gels were run at 120 V for 90 min at 4 °C and then imaged with a Typhoon 9400 scanner using Cy5 settings. Images were quantified by ImageQuant 5.2 software, and data were fitted to the Hill equation to determine the \( K_d \) values using Igor software.

EMSA was carried out with fluorescein-labeled minimal NELFapt as described elsewhere.

**RNA secondary-structure predictions.** The average absolute effect of all three mutations at each position was calculated using the following equation: \[ \frac{\log_2|K_d\text{-mut1}/K_d\text{-wt}| + \log_2|K_d\text{-mut2}/K_d\text{-wt}| + \log_2|K_d\text{-mut3}/K_d\text{-wt}|}{(\text{number of point mutants observed at that position})}. \]
We used Kinefold41 to predict the folded structures and the associated folding free energies. KineFold, unlike other programs that predict secondary structures on the basis of minimal free energy, predicts the structure of RNA as it is being synthesized and thus recapitulates what is happening in HiTS-RAP, where the RNA folds cotranscriptionally. This predicted structure is consistent with the published secondary structure of the GFP aptamer that was supported by extensive mutational analysis21.

Predicted secondary structures and the Gibbs free energy ($\Delta G$) of the folded structures are obtained from KineFold, unless indicated otherwise; however, the actual drawings are obtained from mFold42 because they are easier to manipulate.

37. Sheffield, P., Garrard, S. & Derewenda, Z. Overcoming expression and purification problems of RhoGDI using a family of “parallel” expression vectors. Protein Expr. Purif. 15, 34–39 (1999).
38. McCullum, E.O., Williams, B.A.R., Zhang, J. & Chaput, J.C. Random mutagenesis by error-prone PCR. Methods Mol. Biol. 634, 103–109 (2010).
39. Nielsen, R., Paul, J.S., Albrechtsen, A. & Song, Y.S. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12, 443–451 (2011).
40. Barlow, R. & Blake, J.F. Hill coefficients and the logistic equation. Trends Pharmacol. Sci. 10, 440–441 (1989).
41. Xayaphoummine, A., Bucher, T. & Isambert, H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. Nucleic Acids Res. 33, W605–W610 (2005).
42. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415 (2003).