

1 **Scans of the *MYC* mRNA reveal multiple stable secondary
2 structures—including a 3' UTR motif, conserved across
3 vertebrates, that can affect gene expression.**

4 Collin A. O'Leary^{1,†}, Ryan J. Andrews^{1,†}, Van S. Tompkins¹, Jonathan L. Chen², Jessica L.
5 Childs-DiNEY², Matthew D. Disney², and Walter N. Moss^{1,*}

6 ¹Roy J. Carver Department of Biophysics, Biochemistry and Molecular Biology, Iowa State
7 University, Ames, IA 50011, United States of America

8 ²Department of Chemistry, The Scripps Research Institute, Jupiter, FL 33458, United States of
9 America.

10 [†]These authors contributed equally to this work

11 ***Corresponding Author:** wmooss@iastate.edu

12 Short Title: MYC structure scan

13 **Keywords**

14 MYC, UTR, post-transcriptional regulation, RNA, miRNA

15 **Abstract**

16 The *MYC* gene encodes a human transcription factor and proto-oncogene that is dysregulated
17 in over half of all known cancers. To better understand potential post-transcriptional regulatory
18 features affecting *MYC* expression, we analyzed secondary structure in the *MYC* mRNA using a
19 program that is optimized for finding small locally-folded motifs with a high propensity for
20 function. This was accomplished by calculating folding metrics across the *MYC* sequence using
21 a sliding analysis window and generating unique consensus base pairing models weighted by
22 their lower-than-random predicted folding energy. A series of 30 motifs were identified, primarily
23 in the 5' and 3' untranslated regions, which show evidence of structural conservation and
24 compensating mutations across vertebrate *MYC* homologs. This analysis was able to
25 recapitulate known elements found within an internal ribosomal entry site, as well as discover a
26 novel element in the 3' UTR that is unusually stable and conserved. This novel motif was shown

27 to affect *MYC* expression: likely via modulation of miRNA target accessibility. In addition to
28 providing basic insights into mechanisms that regulate *MYC* expression, this study provides
29 numerous, potentially druggable RNA targets for the *MYC* gene, which is considered
30 “undruggable” at the protein level.

31 **Introduction**

32 The *MYC* proto-oncogene is an important transcription factor that is required for programmed
33 cell death (apoptosis) and cell proliferation [1]. It is a key component of oncogenesis [2] and,
34 indeed, *MYC* is dysregulated in >50% of all cancers [3]. Post-transcriptional control plays
35 significant roles in the regulation of many genes including *MYC*. Within the 5' untranslated
36 region (UTR) of the *MYC* mRNA is a structured internal ribosomal entry site (IRES) that
37 stimulates cap-independent translation under conditions where cap-dependent translation is
38 inhibited: e.g. during apoptosis [4]. Consistent with other IRESs [5] the *MYC* IRES secondary
39 structure, deduced from *in vitro* chemical probing data [6], is complex and contains two
40 pseudoknots—motifs comprised of “non-nested” base pairing between looped out regions of
41 RNA [7]. In addition to the IRES, other post-transcriptional regulatory mechanisms affect *MYC*
42 expression—e.g. microRNAs (miRs) [8]—that may be affected by RNA structure [9].

43 To determine if other structured RNA regulatory elements can be playing roles in *MYC*
44 expression, we applied a methodological pipeline for RNA motif discovery that was optimized
45 from studies of the *Xist* lncRNA [10], as well as the Human [11], Zika and HIV genomes [12].
46 There are two major steps in this pipeline: (1) a scanning step, where the RNA is examined
47 using a sliding analysis window to record predicted metrics important for analyzing RNA
48 secondary structure (e.g. the thermodynamic stability); and, (2) an analysis step where unique
49 local motifs are defined then evaluated vs. comparative sequence/structure and/or experimental
50 probing data. Each step is achieved using the programs ScanFold-Scan and ScanFold-

51 Fold, respectively. Used together these programs define the potential RNA structural properties
52 of long sequences and identify motifs likely to be ordered to form, presumably functional,
53 defined structures. This is accomplished by generating consensus structure models across all
54 scanning windows, where base pairs are weighted by their thermodynamic z-score: a measure
55 of the unusual stability of a sequence that is calculated by comparison to the folding energy of
56 matched randomized control sequences. Here, negative values indicate sequences that are
57 ordered to fold and that may be functional [13]. While the primary goal is to deduce what
58 nucleotides may be functionally significant, ScanFold-Fold models can also increase
59 prediction accuracy [12]

60 In this report, ScanFold-Scan and ScanFold-Fold were applied to the longest MYC RefSeq
61 mRNA isoform to generate a map of its folding landscape as well as deduce motifs important to
62 the regulation of expression. Numerous motifs were deduced, including those that recapitulated
63 known structures in the MYC IRES.

64 **Results**

65 ScanFold-Scan *mapping of secondary structure in the MYC mRNA*

66 To predict RNA secondary structural characteristics important to MYC function, the RefSeq
67 mRNA (NM_001354870) was analyzed using the program ScanFold-Scan [12]. This isoform
68 was selected for analysis, as it would contain all potential structural elements found in other
69 (shorter) MYC isoforms. The mRNA sequence was analyzed using a 1 nt step and 70 nt window
70 size (Figure 1 and Document S1). Several folding metrics were calculated across analysis
71 windows, which are described in detail in the Materials and Methods and in reference [12].
72 Briefly, the ΔG° measures the minimum (lowest or most stable) predicted change in the Gibb's
73 free energy upon RNA folding and indicates the thermodynamic stability of RNA structure. The
74 ensemble diversity (ED) is a measure of the structural diversity predicted in the folding

75 ensemble: low numbers indicate one or few dominant structures, while higher numbers indicate
76 multiple conformations or a lack of structure. The z-score measures the propensity of a
77 sequence to be ordered to fold into stable structures. Negative z-scores give the number of
78 standard deviations more thermodynamically stable a sequence is vs. random (see Eq. 1).

79 The global trends in each metric are shown at the top of Figure 1. The trend in the predicted
80 thermodynamic stability approximately follows the GC% and decreases across *MYC*: going from
81 the highly stable 5' UTR to the relatively unstable 3' UTR. This trend is also discernible in the
82 distributions of predicted Δ Gs for windows spanning the 5' UTR, coding region and 3' UTR of
83 the mRNA (Fig. S1). The windows spanning the 5' and 3' UTR junctions are less stable than
84 flanking sequences. ED values are more evenly distributed across *MYC*, however, jumps in
85 values for windows spanning the 5' and 3' UTR coding region junctions were observed (Figs. 1
86 and S1). Thermodynamic z-scores ranged from highly negative (-5.0; or 5 standard deviations
87 more stable than random) values to positive ones (+2.8). The average z-score across *MYC* was
88 only slightly negative (-0.4) and there was no evidence of global bias in z-score toward negative
89 values. Notably, windows spanning the 5' UTR junction were shifted toward *positive* z-scores
90 (average of +1.4). These trends in RNA structural stability are more striking when considering
91 "short" UTR isoforms for *MYC* (Fig. S2), which end just upstream of the *MYC* IRES and
92 downstream of Motif 17 (Fig. 1).

93 *ScanFold-Fold prediction of functional RNA structural motifs*

94 To deduce local RNA folding that may be functionally significant, all ScanFold-Scan prediction
95 windows were analyzed using ScanFold-Fold. The ScanFold-Fold program generates
96 weighted consensus secondary structures, where minimum free energy (MFE) base pairs that
97 contribute to low z-scores are deduced across the scans. Using a cutoff of -1 ScanFold-Fold
98 identified 354 bp (Document S2) across the mRNA, while a cutoff of -2 yields 46 bp that are

99 localized to the 3' UTR. Refolding the mRNA with -1 ScanFold-Fold bp as constraints added
100 153 bp to the discovered motifs: e.g. by extending helices or closing unpaired bases in the
101 consensus prediction (Document S2). These 507 bp are divided into 30 motifs that span the
102 MYC mRNA (Fig. 1). Motif locations, as expected, correspond to negative dips in z-score;
103 however, dips in ΔG° and ED are also observed at motif sites. The most prominent regions with
104 dips in metrics occur at Motifs 17 and 18 (Fig. 1), which contain very low z-score base pairs
105 (cutoff < -2) deduced by ScanFold-Fold. These two motifs, particularly Motif 17, had the most
106 favorable ScanFold metrics of any region/motif predicted for MYC.

107 All motif bp were analyzed versus an alignment of 15 vertebrate mRNA sequences (Document
108 S3). Motif 17 had the highest conservation of structure and was supported by the greatest
109 number of consistent and compensatory mutations (Fig. 1). In general, Motifs 7–19 showed
110 evidence of conservation, however, little conservation data was found outside these regions:
111 particularly downstream of Motif 19, where the long 3' UTR annotated for the human MYC
112 RefSeq mRNA is not present in the RefSeq mRNA annotations of other species (Document S3).

113 *Analysis of the MYC 5' and 3' UTRs*

114 Motifs 8 and 9 overlap a previously-studied structural feature of the MYC mRNA: the IRES [4,
115 6]. Motif 8 is recapitulated in the MYC IRES structure Domain 1; only the base pairs in the
116 hairpin spanning nt 110 to 136 (Fig. S3) are shifted over to allow the formation of pseudoknot
117 helix α . Motif 9 partially overlaps Domain 2, where nt 284 to 299 of Domain 2 are refolded into
118 two hairpins (Fig. S3). Structure models were compared vs. an alignment of 50 vertebrate MYC
119 UTR sequences (Document S4). The alternative models for Doman 2/Motif 9 are roughly
120 equally well supported by comparative data. Both are comprised of base pairs conserved across
121 vertebrates and that show evidence of possible compensatory mutations: e.g. C279–G284 in
122 Domain 2 vs. A307–U334 and G309–C332 in Motif 9 (Fig. S3). Neither model can be discarded
123 based on these data. Nucleotides within Motif 9 were found to be highly reactive to chemicals in

124 the previous *in vitro* analysis of the *MYC* UTR [6], thus their modeling as single stranded RNA.
125 When overlaid on Motif 9, however, only 4 out of 21 modification sites were inconsistent with the
126 ScanFold-Fold generated model (Fig. S3); additionally, sites of AMV reverse transcriptase
127 pausing suggest that this region is structured.

128 Across the *MYC* mRNA, predicted structural metrics are most favorable in the windows that
129 overlap Motif 17 in the 3' UTR (Fig. 1). There are marked dips in the ΔG° , ED and z-score; all
130 indicating importance of structure in this region. The ScanFold-Fold predicted base pairs in
131 Motif 17 are also the best-conserved across the 15 vertebrate alignment. Previous work on
132 post-transcriptional regulation of *MYC* found that inclusion of the short 3' UTR sequence led to
133 repression of luciferase expression [14] due to the inclusion of a miR-34 binding site. To
134 determine if RNA structural features in the short 3' UTR (beyond Motif 17) could be playing
135 additional roles, the entire sequence was refolded while constraining Motif 17 base pairs. The
136 resulting global short UTR model (Fig. 2) places the ScanFold-Fold predicted Motif 17 into a
137 multibranch loop structure that includes a novel short hairpin. An additional hairpin is also
138 predicted downstream of the multibranch loop. The short 3' UTR model was analyzed vs. an
139 alignment of 59 vertebrate *MYC* 3' UTR sequences (Document S5). This found the highest
140 levels of base pair conservation in the two long Motif 17 hairpins (92% conservation), while the
141 remaining structures are not well-conserved (64% conservation of base pairing). When
142 mutations occur in the highly-conserved Motif 17 they preserve base pairing: e.g. four
143 compensatory (double point) mutations are found in each hairpin in addition to four and two
144 consistent (single point) mutations, respectively (Fig. 2). To see if an orthogonal approach would
145 confirm the 3' UTR model structure or, perhaps, yield a better-conserved alternative model, the
146 program RNAalifold [15] was used to evaluate the short 3' UTR alignment without any base
147 pairing constraints. The RNAalifold program considers both the folding energy and comparative
148 sequence data (implicitly) in prediction; the resulting consensus model (Figure S5) predicts

149 conserved structures that correspond to the two highly-conserved Motif 17 hairpins predicted by
150 ScanFold-Fold.

151 *Functional analyses of the MYC 3' UTR*

152 As the most significant motifs predicted in *MYC* occurred in the 3' UTR, a known site of miRNA
153 targeting, the locations of *MYC*-targeting miRNA binding sites were queried vs. predictions of
154 structure. Of nine miRNAs with known interaction sites [8, 14, 16-20], *seven* occurred within
155 Motif 17 (Fig. 3A and B). miR-34a/b/c, miR-449c and let-7a have overlapping seed binding sites
156 in the unstructured region between the two highly-conserved Motif 17 hairpins (Fig. 3A and B).
157 miR-145 binds downstream and partially overlaps the second hairpin. miR-148 has a seed
158 binding site on the terminal stem-loop of the second hairpin. Interestingly, the two miRNAs that
159 bind outside Motif 17 also do so in other ScanFold-Fold predicted structural motifs: miR-24
160 binds in the stem region of Motif 18 (Fig. 3C), while miR-185 binds toward the 5' end of Motif 15
161 (Fig. 3D). In all cases, conserved RNA structures are predicted to partially occlude miRNA
162 target binding, potentially modulating their effects.

163 Motif 17 was selected for additional experimental analysis due to it having the strongest
164 ScanFold prediction metrics (Fig. 1), high level of structure conservation (Fig. 2), and the
165 presence of multiple miRNA binding sites (Fig. 3A and B). To assess the potential gene
166 regulatory roles of this motif, a luciferase reporter construct was generated incorporating Motif
167 17, along with 27 nt upstream and 11 nt downstream (including a poly(U) tract; Fig. 2). This
168 sequence was inserted into the 3' UTR of the *Renilla luciferase* (RL) expressing pIS2 vector
169 (referred to as the pIS2-M17 [Motif 17] vector; detailed in the Experimental Procedures). When
170 assayed, the pIS2-M17 vector showed a significant decrease in both relative response ratio
171 (RRR; Fig. 4)—a measure of luciferase activity—and translational efficiency (TE) when
172 compared to the unregulated pIS2 control: a ~24% and ~68% decrease in RRR and TE,
173 respectively. This is consistent with previous analyses of the MYC 3' UTR, where the entire

174 short UTR isoform was incorporated into pLSV (an analogous Luciferase vector) and, using a
175 similar analysis pipeline, was shown to lead to gene repression [14]. Similarly, ablation of the
176 miR-34a-c seed (and also, seed regions for miRs 449c and let-7a) showed that miRNA targeting
177 was responsible for the repressive effects of this region.

178 To determine if RNA structure present in Motif 17 influences miRNA binding/repression, two
179 mutant constructs, pIS2-AS1 (ablate stem 1) and pIS2-LS1 (lock stem 1), were designed to
180 increase or decrease miRNA site accessibility respectively (Fig. 4) according to the $\Delta\Delta G$ metric
181 of Kurtesz et. al [21]. This metric accounts for both the energy needed to break native mRNA
182 secondary structure and the energy gained by miRNA binding, and it was used to predict
183 miRNA site accessibility for the WT and mutant constructs. The WT sequence, pIS2-M17, has a
184 predicted $\Delta\Delta G$ of -4.67, whereas pIS2-AS1 and pIS2-LS1 have values of -13.56 (more
185 accessible) and +2.55 (less accessible) respectively. When assayed, pIS2-AS1 shows (Fig. 4) a
186 ~11% decrease in its RRR and an increase in TE of ~20% compared to pIS2-M17, however,
187 these values are not significantly different from each other (Fig. 4). pIS2-LS1 showed a
188 substantial increase in both RRR and TE (~68% and ~70% increase respectively) when
189 compared to pIS2-M17.

190 **Discussion**

191 The analyses performed in this report provide insights into the functions of RNA secondary
192 structure in expression of *MYC*. The ScanFold-Scan results map out global features of RNA
193 structure across the *MYC* mRNA. Interesting trends are observed moving across the sequence,
194 where RNA structure thermodynamic stability decreases going 5' to 3', with marked "jumps" in
195 instability observed at the UTR/coding-region junctions (Figs. S1 and S2). Likewise shifts toward
196 more positive ED and z-score values were observed in junction-spanning windows: most
197 dramatically at the 5' junction, which includes both the CUG (non-canonical) and AUG

198 (canonical) translation initiation sites. These results indicate a lack of stable structure here,
199 reiterating previous observations that indicate inhibitory roles for thermodynamically stable RNA
200 secondary structure at initiation sites [22]. We additionally find evidence that evolution may be
201 specifically selecting for *MYC* initiation site sequences that are ordered to be less-stable than
202 that predicted for sequences of similar composition (thus the positive z-scores); as well, the
203 junction sequence is expected to have a volatile conformational ensemble, where no particular
204 structure dominates (high ED).

205 The high and low respective thermodynamic stabilities of the 5' and 3' UTRs (Figs. S1 and S2)
206 indicate differing roles for RNA folding in these regions. The highly-stable 5' UTR would be
207 expected to inhibit canonical translation by obstructing scanning ribosomes; thus, the presence
208 of an IRES in the *MYC* mRNA. This can provide mechanisms for fine-tuning the post-
209 transcriptional regulation of the *MYC* gene by allowing it to be translated in a cap-independent
210 manner. The *MYC* IRES was shown to be active in some, but not all tissue types and the
211 variability of activity is attributed to the presence, or lack of, trans-regulatory elements (e.g.
212 RBPs; [4]). This demonstrates how cis-elements of the mRNA can interact with trans-regulatory
213 elements to diversify (i.e. regulate) the cellular levels of a protein.

214 In contrast, the low stability of the 3' UTR suggests a need for increased accessibility of the
215 mRNA sequence: e.g. for intermolecular interactions with post-transcriptional regulatory factors
216 such as miRNAs and regulatory proteins. Counterintuitively, the sites with the greatest evidence
217 of having been ordered to fold into a specific structure are in the 3' UTR (e.g. Motifs 17 and 18
218 in Fig. 1). Motif 17, for example, is the most well-conserved structured region in *MYC*—even
219 more so than the IRES domain (Figs. 2 and S3)—and is supported via multiple compensatory
220 and consistent base mutations. The highly favorable metrics and deep conservation of this motif
221 throughout vertebrates indicated its biological importance, which was borne out by the analysis
222 of miRNA binding sites (Fig. 3A and B) and Motif 17 function (Fig. 4). This motif acts as a hub

223 for miRNA interactions and may organize the *MYC* miRNA target sites for interactions: e.g. the
224 two highly-conserved hairpins may act as a “structure cassette” for maintaining the single-
225 strandedness of the miR-34, -449c, and -let-7a seed binding regions, while modulating the
226 accessibility of additional bases for miRNA pairing that can affect the outcome of miRNA
227 targeting: e.g. translational repression.

228 The second most favorable motif (Motif 18; Fig. 1) contains a miR-24 interacting region (Fig.
229 3C). Notably, this interaction is “seedless” [17]—only three of the miR-24 seed nt are base
230 paired to *MYC* (Fig. 3C). Most of the miR-24-interacting nt on *MYC* are predicted to be bound
231 up in structure. Here, as in other interaction sites, RNA folding may be modulating accessibility.
232 We observed that reducing the accessibility of miRNA binding in Motif 17 via pIS2-LS1
233 increased RRR and TE—potentially by reducing the amounts of miRNA-mediated gene
234 repression. On the other hand, pIS2-M17 (the WT motif) and pIS2-AS1 (that ablates the first
235 stem loop) had similar RRR and TE. Additional cellular factors may influence how these
236 structures affect miRNA targeting: e.g. interactions with RNA binding proteins or post-
237 transcriptional modifications can affect RNA folding. Thus, structural motifs and their additional
238 interactions and alterations may be a way to fine tune the effects of miRNA regulation of *MYC*.

239 Conserved RNA structural motifs may also serve other functions in regulating *MYC* expression.
240 Both Motif 17 and 18 occur in a particularly GC-poor and thermodynamically unstable region of
241 the 3' UTR (Fig. 1). This is due to long stretches of highly-conserved poly(U) (uridine) tracts that
242 occur between and around these structural motifs (Documents S3 and S5). Interestingly,
243 poly(U) tracts have been found to stabilize mRNA sequences via structural interactions with the
244 poly(A) (adenosine) tail [23]. Thus, an additional function of *MYC* 3' UTR structural motifs may
245 be to organize poly(U) tracts to facilitate interactions with the poly(A) tail. For example, in Figure
246 2 the poly(U) sequence is bulged out between the second Motif 17 hairpin and the basal stem of
247 the predicted multibranch loop, which could allow the poly(A) tail to “wind” around this structure:

248 e.g. in an analogous way to poly(U)-poly(A) interactions in viral and ncRNA stabilizing elements
249 [24, 25]. Notably, our qPCR data does not indicate degradation of miRNA targeted RL
250 transcripts. Instead, slight transcript accumulation is observed in experimental samples (which
251 contain the poly(U) tract) compared to unregulated pIS2 samples (Table S1). Motif 17, in
252 addition to regulating miRNA target accessibility, may be affecting transcript stability by
253 modulating interactions of the poly(U) tract with the poly(A) tail of the mRNA. More investigation
254 is needed to parse out these interactions.

255 Additional functional motifs are predicted beyond Motif 18 (Fig. 1), which may also be
256 functionally significant. Interestingly, cancer-associated *MYC* translocations [26] can lead to
257 UTR truncations that delete predicted motifs: potentially impacting function and contributing to
258 *MYC* dysregulation. Likewise, seven predicted motifs fall within the *MYC* coding region, which
259 may be functionally significant: e.g. by providing roadblocks for translation that can affect protein
260 folding [27] or by affecting interactions with regulatory factors [28]. Notably, miR-185 targets a
261 sequence that overlaps Motif 15 (Fig. 3D), which falls within the *MYC* coding region (Fig. 1).
262 Awareness of the importance of miRNA targeting in coding regions is growing [29] and,
263 presumably, additional *MYC* miRNA interactors remain to be discovered.

264 To conclude, this report provides valuable information on *MYC* mRNA secondary structure that
265 has implications toward a better understanding of post-transcriptional gene regulation. In
266 addition to providing global structural data, discreet local motifs with a high propensity for
267 function are proposed, including a particularly interesting motif in the 3' UTR that has been
268 functionally validated. We showed Motif 17 possesses post-transcriptional regulatory function
269 and, at the very least, this function is a result of structure regulated miRNA targeting. Our
270 findings illustrate the utility of ScanFold-Scan and ScanFold-Fold in finding structured,
271 regulatory motifs and highlight the important role of RNA secondary structure in the post-
272 transcriptional gene regulation of *MYC* expression. This study provides a roadmap for further

273 analyses of the structure/function relationships in the *MYC* mRNA and a framework for
274 understanding other experimental results. For example, identified clinically significant sequence
275 variants can be cross-referenced to these results to deduce their potential impact on RNA
276 folding. Additionally, these results generate a large list of structural motifs that may be
277 druggable targets [30, 31] for *MYC*, which is considered undruggable at the protein level [3].

278 **Materials and Methods**

279 *In silico* analyses

280 The *Homo sapiens* *MYC* RefSeq mRNA sequence was downloaded from the NCBI nt database
281 (GenBank Accession: NM_002467.5). ScanFold-Scan was run using a single nt step size and
282 window sizes of 70 (Document S1) and 120 nt (results with the longer window size were
283 unchanged [data not shown], this 70 nt window was used in subsequent analyses). RNA
284 structural metrics were calculated for windows using the RNAfold algorithm [32] using the
285 Turner energy model [33, 34] at 37 °C. Z-score calculations were performed using the following
286 equation (adapted from the approach of [13]):

287
$$z\text{-score} = (\Delta G^\circ_{native} - \overline{\Delta G^\circ}_{random})/\sigma \quad (\text{Eq. 1})$$

288 Here, ΔG°_{native} is the native sequence minimum free energy (MFE) predicted by RNAfold.
289 $\overline{\Delta G^\circ}_{random}$ is the average MFE predicted for 100X mononucleotide randomized sequences. The
290 standard deviation, σ , is calculated across all sequences. The other calculated data are: the P-
291 value, which measures the fraction of random sequences that are more stable than native in the
292 z-score calculation (this acts as a quality control measure for the z-score); the MFE ΔG° , which
293 measures the thermodynamic stability of RNA secondary structure formation; the MFE base
294 pairs that generate the MFE ΔG° , which are output in “dot-bracket” notation; the ensemble
295 diversity (ED), which provides an estimate of the structural diversity in the RNA conformational

296 ensemble (e.g. low ED indicates a single dominant conformation); the fraction of the (f)MFE in
297 the ensemble, which estimates the contribution of the MFE conformation to the ensemble; the
298 ensemble centroid structure, which is the conformation most similar to others in the ensemble;
299 and the nt frequencies and GC percentages.

300 ScanFold-Scan prediction windows were next analyzed using the program ScanFold-Fold
301 to deduce consensus motifs weighted by the z-score. The ScanFold-Fold method is detailed
302 in [12]. Resulting output consisted of a list of all base pairing partners predicted for each
303 nucleotide of the *MYC* mRNA (Document S6) and a list of the most favorable base pairing
304 partners when weighting by z-score (Document S7). From the latter, base pairs which
305 contributed to consistently negative z-scores (i.e. bps with average z-scores less -1 from
306 Document S7) were used as constraints in an RNAfold prediction on the entire mRNA under the
307 additional constraint of a maximum bp distance of 300 nt. Base pairs that extended ScanFold-
308 Fold helices were identified and used to generate the final motif models (Document S2). For
309 visualizing the results of modeling, 2D rendering were generated using VARNA [35] and figures
310 were produced with Adobe Illustrator.

311 For the analysis of conservation of ScanFold-Fold motifs across *MYC*, homologous mRNAs
312 for 14 representative vertebrates were obtained from the NCBI RefSeq RNA database [36]. This
313 database was also queried using BLAST [37] to deduce homologs for the “short” *MYC* 5' and 3'
314 UTR sequences. Alignments for the mRNA (Document S3) and UTRs (Documents S4 and S5)
315 were performed using MAFFT [38], implementing the MAFFT-E-INS-i and MAFFT-G-INS-i
316 strategies, respectively [39].

317 A global model for the short 3' UTR (defined/used in a previous study of miRNA targeting [14]
318 was generated by constraining base pairs from Motif 17 and refolding the remaining sequence

319 using RNAfold [32]. A consensus secondary structure for the short *MYC* 3' UTR was predicted
320 (Fig. S3) using RNAalifold [15] with the 3' UTR alignment (Document S5) as input.

321 *Experimental analyses*

322 **Cell Culture.** HeLa cells were incubated at 37°C and 5% CO₂ and maintained in DMEM
323 supplemented with 10% FBS, penicillin and streptomycin, and L-glutamine. Cells were
324 passaged at 60-80% confluence and used between 3-40 passages.

325 **Luciferase Vectors.** For our experiments, two luciferase plasmid vector backbones were used.
326 Both the transfection control vector, pIS0, which encoded firefly (FF) luciferase, and the
327 experimental vector, pIS2, which encoded renilla (RL) luciferase, were gifts from David Bartel
328 (Addgene plasmid # 12178 ; <http://n2t.net/addgene:12178> ; RRID:Addgene_12178) and
329 (Addgene plasmid # 12177 ; <http://n2t.net/addgene:12177> ; RRID:Addgene_12177). The
330 pcDNA3.1-miR34a vector was a gift from Heidi Schwarzenbach (Addgene plasmid # 78125 ;
331 <http://n2t.net/addgene:78125> ; RRID:Addgene_78125).

332 To test the post-transcriptional regulation of ScanFold-Fold predicted motifs, the Motif 17
333 sequence, along with 27 nt upstream and 11 nt downstream were incorporated into the 3'-UTR
334 of pIS2 to generate pIS2-M17. Mutants that destabilize, pIS2-AS1, or stabilize, pIS2-LS1, the
335 structure present in pIS2-M17 were generated. For pIS2-AS1, 6 mutations were incorporated
336 that disrupt canonical base pairing in the first conserved hairpin. To generate pIS2-LS1, 3
337 mutations and 1 base deletion were introduced in the bulge on the upstream side of the first
338 conserved hairpin. Mutations that destabilize or stabilize Motif 17 were predicted using the ΔΔG
339 metric as a measure of miRNA site accessibility (Kertesz et al., 2007). 70 nt upstream and 70 nt
340 downstream of the miRNA target site were included in our ΔΔG calculations.

341 The sequences for pIS2-M17, pIS2-AS1, and pIS2-LS1 were ordered as gBlocks from IDT and
342 cloned using AgeI (5') and SpeI (3') restriction sites (sequences in Supplemental Table S2).

343 Insertion of experimental sequences into the 3'-UTR of pIS2 required double restriction enzyme
344 digest (using AgeI and Spel from NEB) of both the gBlock and pIS2, following digestion,
345 fragment and vector DNA were purified (Zymo DNA Clean and Concentrator kit), ligated (T4
346 Ligase from ThermoFischer), and transformed into DH5 α -T1 competent cells using standard
347 procedures. Carbenicillin selected colonies were cultured and plasmids were extracted (Qiaprep
348 kit) and sequenced using an Applied Biosystems 3730xl DNA Analyzer.

349 **Dual Luciferase Assay.** Dual luciferase assays followed recommendations of an established
350 method (Etten et al., 2013). In brief, the pIS0 vector (FF) is transfected at constant levels across
351 all samples to serve as an internal control to which RL luciferase expression is normalized. All
352 samples were run as biological triplicates. HeLa cells were counted using a hemocytometer and
353 plated in a 24-well dish at a density of 50,000 cells per well. After 48 hours, cells were
354 transfected using Lipofectamine 3000 (ThermoFischer) with 500 nanograms total plasmid DNA
355 at a 1:1:8 ratio (pIS0:pIS2-based:pcDNA3.1-miR34a). Twenty-four hours later, cells were
356 trypsinized, resuspended, and split into each of a 24-well plate for RNA analysis (1 ml) and a
357 96-well plate for the dual luciferase assay (0.2 ml). After another 24-hour incubation, cells in the
358 96-well dish were lysed, and luciferase activity was measured using Promega's Dual Luciferase
359 Reagent Assay kit on a Biotek Synergy 2 plate reader with a collection time of 10 seconds.
360 Relative response ratios (RRR), the ratio of RL to FF relative light units (RLUs), were calculated
361 for each sample and then normalized to the empty, unregulated pIS2 RRR. Cells from the 24-
362 well plate were placed in TRIzol (ThermoFisher) and either stored at -80°C or immediately
363 processed as below.

364 **RNA Processing and qPCR Analysis.** Cellular RNA was purified from samples in TRIzol using
365 Zymo's Direct-Zol RNA Miniprep kit. Purified RNA was then Dnase I treated (NEB) for 2 hours at
366 37°C and the resulting DNase-treated RNA was purified with Zymo's RNA Clean and

367 Concentrator kit. Reverse transcription was done using 1 microgram of purified RNA, random
368 hexamers, and Superscript III (ThermoFisher).

369 Relative abundance of RL transcripts across samples were measured by qPCR, performed
370 using PowerUp SYBR Green Master Mix on 1% cDNA input on an Applied Biosystems
371 QuantStudio 3 instrument (ThermoFisher). Data were analyzed using the $\Delta\Delta Ct$ method, where
372 the relative abundance of RL transcripts in the samples were determined using the FF transcript
373 as the reference gene. Translational efficiencies (TE), a normalization metric ($RRR/2^{[-\Delta\Delta Ct_{RL}]}$),
374 were calculated for each sample. Primers used in qPCR were: RL FWD 5'-
375 ggaattataatgcttatctacgtgc-3'; RL REV 5'-cttgcgaaaaatgaagacccttac-3'; FF FWD 5'-
376 ctcactgagactacatcagc-3'; and FF REV 5'-tccagatccacaacccttcgc-3'.

377 All data are available in the Supplemental Information. ScanFold-Scan and ScanFold-Fold
378 are available for download from GitHub: <https://github.com/moss-lab/ScanFold>. RNAfold and
379 RNAalifold are both bundled within the ViennaRNA package [32]: available at:
380 <https://www.tbi.univie.ac.at/RNA/>.

381 Acknowledgments

382 This research was supported by NIH/NIGMS grant R00GM112877 (WNM) and startup funds
383 from the Iowa State University College of Agriculture and Life Sciences and the Roy J. Carver
384 Charitable Trust (WNM) as well as R01-GM097455 (MDD).

385 Figure Legends

386 **Figure 1.** Summary of ScanFold-Scan and ScanFold-Fold results for the short UTR MYC
387 mRNA. At the top are charts indicating the predicted ScanFold-Scan metrics across the
388 mRNA. The bars are set at the 1st nt of the 70 nt window, thus data corresponds to the 70 nt
389 downstream of the bar. Below these is a cartoon of the MYC mRNA with UTRs and coding

390 region represented in thin and thick black lines, respectively. This cartoon is annotated with
391 boxes which depict the location and extent of ScanFold-Fold predicted motifs shaded red
392 based on the average z-score of windows in which motif base pairs occurred. Below these are
393 RNA secondary structure arc diagrams which depict the most favorable base pairs predicted via
394 ScanFold-Fold, colored according the average z-scores of windows in which they appear
395 (with blue, green and yellow corresponding to less than -2, -1 and 0 z-scores averages
396 respectively). Below these, are refolded models of the motifs built with -1 average z-score bps
397 as constraints. Each is annotated with their bp conservation as determined from an alignment of
398 15 representative mRNAs (Document S3) indicated by shading on the base pair (see key).
399 Circled bases are sites of putative structure-preserving consistent and compensatory mutations.

400 **Figure 2.** Short MYC 3' UTR model. HP2–3 comprise Motif 17 of the ScanFold-Fold results,
401 which were constrained in the calculation. Base pair conservation shading indicated in the key
402 and data are taken from a comparison of 59 RefSeq mRNA vertebrate alignment (Document
403 S5).

404 **Figure 3.** Annotations of miRNA binding sites on ScanFold-Fold predicted motifs. (A) Shows
405 miRNA sequences above the “dot-bracket” structure of Motif 17 (matched brackets indicated
406 base pairs). Seed sites and the complements on Motif 17 are colored. (B) Shows miRNA seed
407 binding sites annotated on the 2D model of Motif 17. (C) Shows base-pairing between miR-24
408 and the 2D model of Motif 18. (C) Shows base-pairing between miR-24 and the 2D model of
409 Motif 15.

410 **Figure 4.** RRRs and TEs were calculated for each set of samples and normalized to the
411 unregulated activity of pIS2. Error bars report the standard error. Experimental vectors pIS2-
412 M17, pIS2-AS1, and pIS2-LS1 all display activity that differs from pIS2. pIS2-M17 and pIS2-AS1
413 both show decreased RRRs while pIS2-LS1, which was designed to have a more stable and
414 less accessible 3'-UTR, shows an increase in RRR compared to pIS2. pIS2-MSC displays a

415 large decrease in TE and pIS2-AS1, which is predicted to have a more accessible miRNA site,
416 displays a TE which is slightly increased compared to pIS2-M17. The TE of pIS2-LS1 is
417 markedly greater than pIS2-M17, possibly reflecting the decrease in miRNA target site
418 accessibility. The $\Delta\Delta G$ values for pIS2-M17, pIS2-AS1, and pIS2-LS1 are also shown along with
419 destabilizing pIS2-AS1 mutations (displayed in red) and stabilizing pIS2-LS1 mutations
420 (displayed in green).

421 **Supplemental Material**

422 Supplemental material is available for this report.

423 **References**

- 424 1. Dang CV. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol
425 Cell Biol.* 1999;19(1):1-11. Epub 1998/12/22. PubMed PMID: 9858526; PubMed Central
426 PMCID: PMCPMC83860.
- 427 2. Dang CV. MYC on the path to cancer. *Cell.* 2012;149(1):22-35. Epub 2012/04/03. doi:
428 10.1016/j.cell.2012.03.003. PubMed PMID: 22464321; PubMed Central PMCID:
429 PMCPMC3345192.
- 430 3. Dang CV, Reddy EP, Shokat KM, Soucek L. Drugging the 'undruggable' cancer targets.
431 *Nat Rev Cancer.* 2017;17(8):502-8. Epub 2017/06/24. doi: 10.1038/nrc.2017.36. PubMed PMID:
432 28643779; PubMed Central PMCID: PMCPMC5945194.
- 433 4. Stoneley M, Subkhankulova T, Le Quesne JP, Coldwell MJ, Jopling CL, Belsham GJ, et
434 al. Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the
435 nuclear compartment. *Nucleic Acids Res.* 2000;28(3):687-94. Epub 2000/01/19. PubMed PMID:
436 10637319; PubMed Central PMCID: PMCPMC102558.
- 437 5. Yamamoto H, Unbehaun A, Spahn CMT. Ribosomal Chamber Music: Toward an
438 Understanding of IRES Mechanisms. *Trends Biochem Sci.* 2017;42(8):655-68. Epub
439 2017/07/08. doi: 10.1016/j.tibs.2017.06.002. PubMed PMID: 28684008.
- 440 6. Le Quesne JP, Stoneley M, Fraser GA, Willis AE. Derivation of a structural model for the
441 c-myc IRES. *J Mol Biol.* 2001;310(1):111-26. Epub 2001/06/23. doi: 10.1006/jmbi.2001.4745.
442 PubMed PMID: 11419940.
- 443 7. Liu B, Mathews DH, Turner DH. RNA pseudoknots: folding and finding. *F1000 Biol Rep.*
444 2010;2:8. doi: 10.3410/B2-8. PubMed PMID: 20495679; PubMed Central PMCID:
445 PMCPMC2873773.
- 446 8. Jackstadt R, Hermeking H. MicroRNAs as regulators and mediators of c-MYC function.
447 *Biochim Biophys Acta.* 2015;1849(5):544-53. Epub 2014/04/15. doi:
448 10.1016/j.bbagr.2014.04.003. PubMed PMID: 24727092.
- 449 9. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome
450 through RNA structure. *Nat Rev Genet.* 2011;12(9):641-55. doi: 10.1038/nrg3049. PubMed
451 PMID: 21850044; PubMed Central PMCID: PMCPMC3858389.
- 452 10. Fang R, Moss WN, Rutenberg-Schoenberg M, Simon MD. Probing Xist RNA Structure in
453 Cells Using Targeted Structure-Seq. *PLoS Genet.* 2015;11(12):e1005668. doi:

454 10.1371/journal.pgen.1005668. PubMed PMID: 26646615; PubMed Central PMCID:
455 PMCPMC4672913.

456 11. Andrews RJ, Baber L, Moss WN. RNAStructuromeDB: A genome-wide database for
457 RNA structural inference. *Sci Rep.* 2017;7(1):17269. doi: 10.1038/s41598-017-17510-y.
458 PubMed PMID: 29222504; PubMed Central PMCID: PMCPMC5722888.

459 12. Andrews R.J. RJ, Moss W.N. ScanFold: an approach for genome-wide discovery of
460 local RNA structural elements—applications to Zika virus and HIV. *PeerJ Preprints.*
461 2018;6:e6136. doi: <https://doi.org/10.7717/peerj.6136>

462 13. Clote P, Ferre F, Kranakis E, Krizanc D. Structural RNA has lower folding energy than
463 random RNA of the same dinucleotide frequency. *RNA.* 2005;11(5):578-91. doi:
464 10.1261/rna.7220505. PubMed PMID: 15840812; PubMed Central PMCID: PMCPMC1370746.

465 14. Kong YW, Cannell IG, de Moor CH, Hill K, Garside PG, Hamilton TL, et al. The
466 mechanism of micro-RNA-mediated translation repression is determined by the promoter of the
467 target gene. *Proc Natl Acad Sci U S A.* 2008;105(26):8866-71. Epub 2008/06/27. doi:
468 10.1073/pnas.0800650105. PubMed PMID: 18579786; PubMed Central PMCID:
469 PMCPMC2449332.

470 15. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved
471 consensus structure prediction for RNA alignments. *BMC Bioinformatics.* 2008;9:474. doi:
472 10.1186/1471-2105-9-474. PubMed PMID: 19014431; PubMed Central PMCID:
473 PMCPMC2621365.

474 16. Luan W, Wang Y, Chen X, Shi Y, Wang J, Zhang J, et al. PKM2 promotes glucose
475 metabolism and cell growth in gliomas through a mechanism involving a let-7a/c-Myc/hnRNPA1
476 feedback loop. *Oncotarget.* 2015;6(15):13006-18. Epub 2015/05/08. doi:
477 10.18633/oncotarget.3514. PubMed PMID: 25948776; PubMed Central PMCID:
478 PMCPMC4536995.

479 17. Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, et al. miR-24 Inhibits cell
480 proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR
481 microRNA recognition elements. *Mol Cell.* 2009;35(5):610-25. Epub 2009/09/15. doi:
482 10.1016/j.molcel.2009.08.020. PubMed PMID: 19748357; PubMed Central PMCID:
483 PMCPMC2757794.

484 18. Sachdeva M, Zhu S, Wu F, Wu H, Walia V, Kumar S, et al. p53 represses c-Myc through
485 induction of the tumor suppressor miR-145. *Proc Natl Acad Sci U S A.* 2009;106(9):3207-12.
486 Epub 2009/02/10. doi: 10.1073/pnas.0808042106. PubMed PMID: 19202062; PubMed Central
487 PMCID: PMCPMC2651330.

488 19. Liao JM, Lu H. Autoregulatory suppression of c-Myc by miR-185-3p. *J Biol Chem.*
489 2011;286(39):33901-9. Epub 2011/08/13. doi: 10.1074/jbc.M111.262030. PubMed PMID:
490 21832077; PubMed Central PMCID: PMCPMC3190790.

491 20. Miao LJ, Huang SF, Sun ZT, Gao ZY, Zhang RX, Liu Y, et al. MiR-449c targets c-Myc
492 and inhibits NSCLC cell progression. *FEBS Lett.* 2013;587(9):1359-65. Epub 2013/03/20. doi:
493 10.1016/j.febslet.2013.03.006. PubMed PMID: 23507140.

494 21. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in
495 microRNA target recognition. *Nat Genet.* 2007;39(10):1278-84. Epub 2007/09/26. doi:
496 10.1038/ng2135. PubMed PMID: 17893677.

497 22. Gu W, Zhou T, Wilke CO. A universal trend of reduced mRNA stability near the
498 translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol.*
499 2010;6(2):e1000664. Epub 2010/02/09. doi: 10.1371/journal.pcbi.1000664. PubMed PMID:
500 20140241; PubMed Central PMCID: PMCPMC2816680.

501 23. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. Global analysis of mRNA isoform
502 half-lives reveals stabilizing and destabilizing elements in yeast. *Cell.* 2014;156(4):812-24. Epub
503 2014/02/18. doi: 10.1016/j.cell.2013.12.026. PubMed PMID: 24529382; PubMed Central
504 PMCID: PMCPMC3939777.

505 24. Tycowski KT, Shu MD, Borah S, Shi M, Steitz JA. Conservation of a triple-helix-forming
506 RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep.*
507 2012;2(1):26-32. Epub 2012/07/31. doi: 10.1016/j.celrep.2012.05.020. PubMed PMID:
508 22840393; PubMed Central PMCID: PMCPMC3430378.

509 25. Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA. Formation of triple-helical
510 structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad
511 Sci U S A.* 2012;109(47):19202-7. Epub 2012/11/07. doi: 10.1073/pnas.1217338109. PubMed
512 PMID: 23129630; PubMed Central PMCID: PMCPMC3511071.

513 26. Ott G, Rosenwald A, Campo E. Understanding MYC-driven aggressive B-cell
514 lymphomas: pathogenesis and classification. *Blood.* 2013;122(24):3884-91. Epub 2013/09/07.
515 doi: 10.1182/blood-2013-05-498329. PubMed PMID: 24009228.

516 27. Faure G, Ogurtsov AY, Shabalina SA, Koonin EV. Role of mRNA structure in the control
517 of protein folding. *Nucleic Acids Res.* 2016;44(22):10898-911. doi: 10.1093/nar/gkw671.
518 PubMed PMID: 27466388; PubMed Central PMCID: PMCPMC5159526.

519 28. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-
520 transcriptional gene regulation. *FEBS Lett.* 2008;582(14):1977-86. Epub 2008/03/18. doi:
521 10.1016/j.febslet.2008.03.004. PubMed PMID: 18342629; PubMed Central PMCID:
522 PMCPMC2858862.

523 29. Zhang K, Zhang X, Cai Z, Zhou J, Cao R, Zhao Y, et al. A novel class of microRNA-
524 recognition elements that function only within open reading frames. *Nat Struct Mol Biol.*
525 2018;25(11):1019-27. Epub 2018/10/10. doi: 10.1038/s41594-018-0136-3. PubMed PMID:
526 30297778; PubMed Central PMCID: PMCPMC6219929.

527 30. Tran T, Disney MD. Identifying the preferred RNA motifs and chemotypes that interact
528 by probing millions of combinations. *Nat Commun.* 2012;3:1125. Epub 2012/10/11. doi:
529 10.1038/ncomms2119. PubMed PMID: 23047683; PubMed Central PMCID: PMCPMC3533436.

530 31. Angelbello AJ, Chen JL, Childs-Disney JL, Zhang P, Wang ZF, Disney MD. Using
531 Genome Sequence to Enable the Design of Medicines and Chemical Probes. *Chem Rev.*
532 2018;118(4):1599-663. Epub 2018/01/13. doi: 10.1021/acs.chemrev.7b00504. PubMed PMID:
533 29322778; PubMed Central PMCID: PMCPMC5989578.

534 32. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al.
535 ViennaRNA Package 2.0. *Algorithms Mol Biol.* 2011;6:26. doi: 10.1186/1748-7188-6-26.
536 PubMed PMID: 22115189; PubMed Central PMCID: PMCPMC3319429.

537 33. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating
538 chemical modification constraints into a dynamic programming algorithm for prediction of RNA
539 secondary structure. *Proc Natl Acad Sci U S A.* 2004;101(19):7287-92. doi:
540 10.1073/pnas.0401799101. PubMed PMID: 15123812; PubMed Central PMCID:
541 PMCPMC409911.

542 34. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for
543 predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* 2010;38(Database
544 issue):D280-2. Epub 2009/11/03. doi: 10.1093/nar/gkp892. PubMed PMID: 19880381; PubMed
545 Central PMCID: PMCPMC2808915.

546 35. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA
547 secondary structure. *Bioinformatics.* 2009;25(15):1974-5. doi: 10.1093/bioinformatics/btp250.
548 PubMed PMID: 19398448; PubMed Central PMCID: PMCPMC2712331.

549 36. O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference
550 sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional
551 annotation. *Nucleic Acids Res.* 2016;44(D1):D733-45. Epub 2015/11/11. doi:
552 10.1093/nar/gkv1189. PubMed PMID: 26553804; PubMed Central PMCID: PMCPMC4702849.

553 37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.
554 *J Mol Biol.* 1990;215(3):403-10. Epub 1990/10/05. doi: 10.1016/S0022-2836(05)80360-2.
555 PubMed PMID: 2231712.

556 38. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple
557 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059-66.
558 Epub 2002/07/24. PubMed PMID: 12136088; PubMed Central PMCID: PMCPMC135756.
559 39. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. *Methods*
560 *Mol Biol.* 2014;1079:131-46. Epub 2013/10/31. doi: 10.1007/978-1-62703-646-7_8. PubMed
561 PMID: 24170399.

562

563

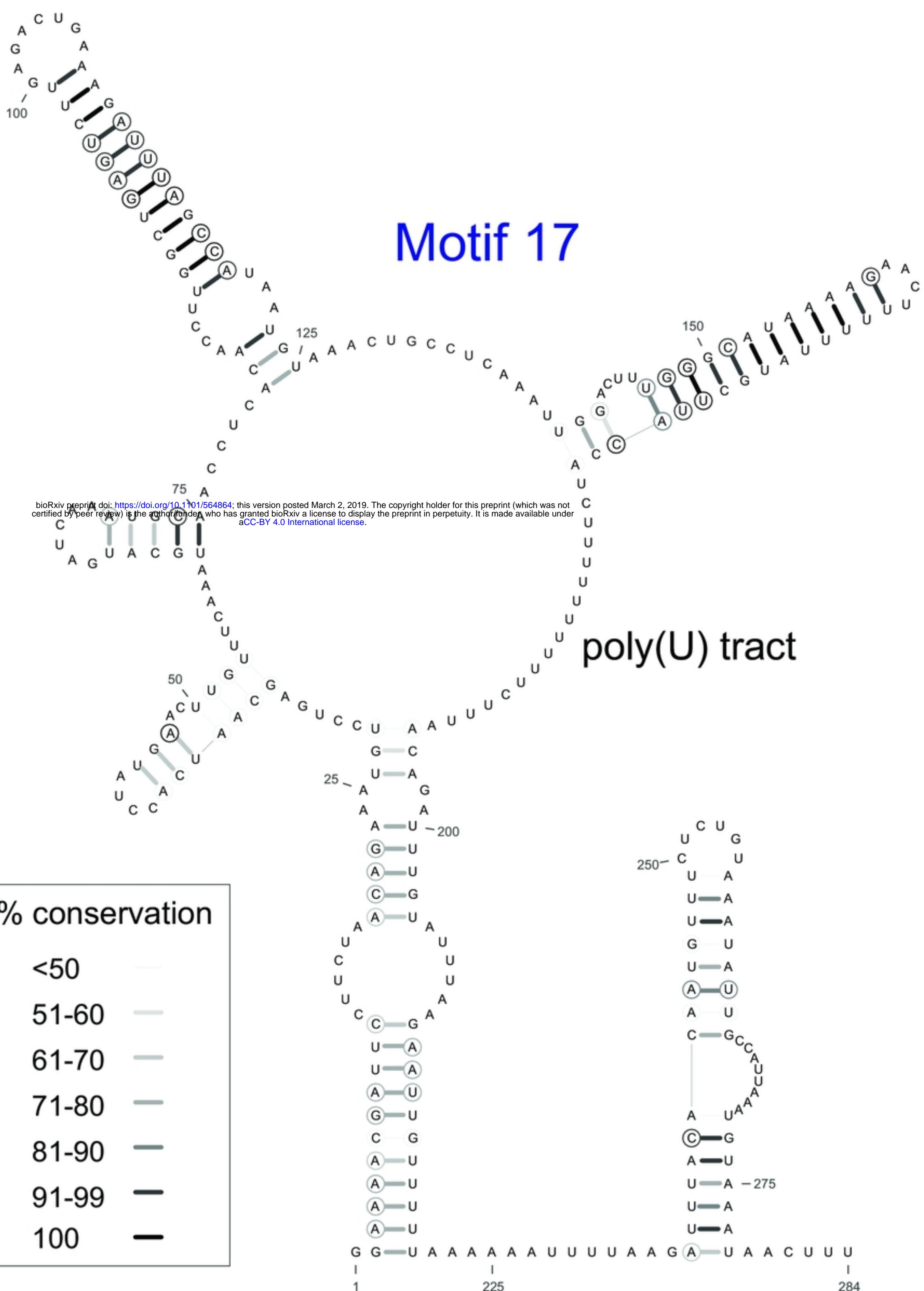


Figure 2

A

miR-145 3'-UUCCCUAAGGACCCUUUUGACCUU-5'
 miR-let-7a 3'-UUGAUUAUGUUGGAU**GAUGGAGU**-5'
 miR-449c 3'-UGUCGGCGAUCGUUAUG**UGACCGGA**U-5'
 miR-34c 3'-CGUUAGUCGAUUGAUG**UGACCGG**A-5'
 miR-34b 3'-GUUAGUCGAUUACUG**UGACCGG**AU-5'
 miR-34a 3'-UUGUUGGUCGAUUCUG**UGACCGG**U-5' 3'-UCAGCCUCACAGAG**UCUUGAA**A-5' miR-148a

5'-ACAACCUUGGCUGAGUCUUGAGACUGAAAGAUUUAGCCAUAUAUGUAA**ACUGCCUCAAAUGG**CUUUGGGCAUAAA**AGAACUU**UUUUUAGCUUACCA-3'
 (((.....((((((.....)))))))).....((.....((((((.....))))))))

B

2616 G A C U
 A G
 G A
 U A
 U A
 C G
 U A
 G U
 A U
 G U
 U A
 C G
 G C
 G C
 U A
 C A
 C A
 A U
 C G
 A G
 A G
 A G
 miR-148a A C
 G U
 G U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 A U
 miR-let-7a C G
 G C
 G U
 G U
 miR-449c G U
 G U
 G U
 miR-34c U A
 U A
 U A
 miR-34b U U A
 U C
 U C
 miR-34a A G C
 G C
 G C
 2712 3'-G A C A A G G A • U - A
 2712 2779 - A U G U U U C 2847
 miR-145 2616

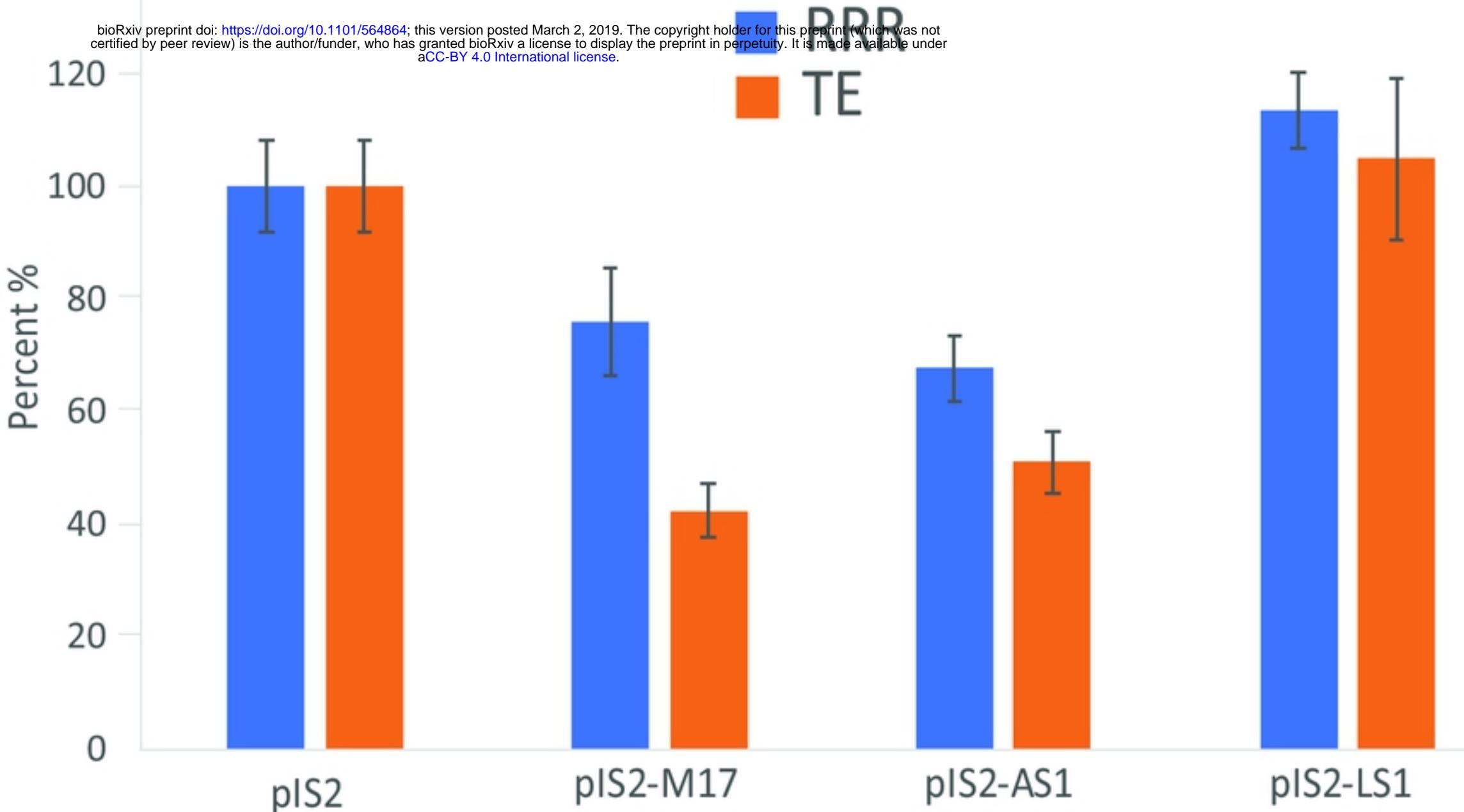
C

5' U - A U
 U • A
 G • C A
 G • C G
 C • G C
 U U - A
 C U - G
 A A - U
 G A
 U A
 U A
 U A
 U A
 U A
 U A
 U A
 miR-24 3'-G A C A A G G A • U - A
 2779 - A U G U U U C 2847

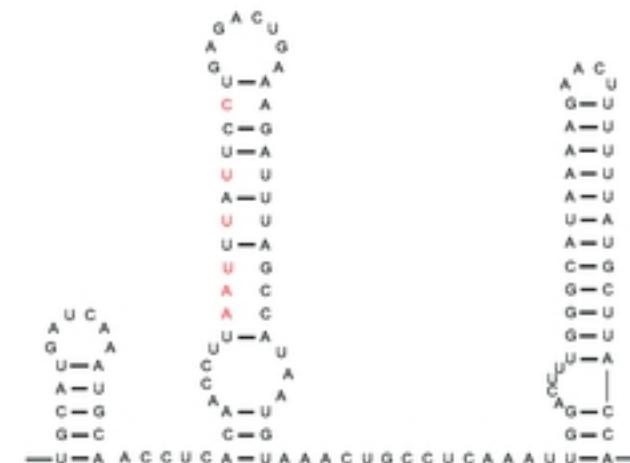
D

5' G G A
 A A - U
 G • C - G U C
 G • C - G A
 G • C - G
 G • C - G
 G • C - G
 2219 - C A A C A A C C G A A A A U G C A C C A 2276

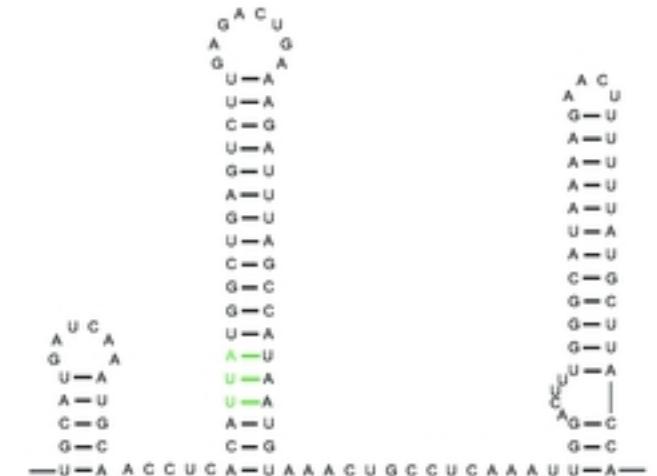
Figure 3



pIS2-M17
 $\Delta\Delta G = -4.67 \text{ kcal/mol}$



plS2-AS1
 $\Delta\Delta G = -13.56 \text{ kcal/mol}$



pIS2-LS1
 $\Delta\Delta G = +2.55 \text{ kcal/mol}$

Figure 4

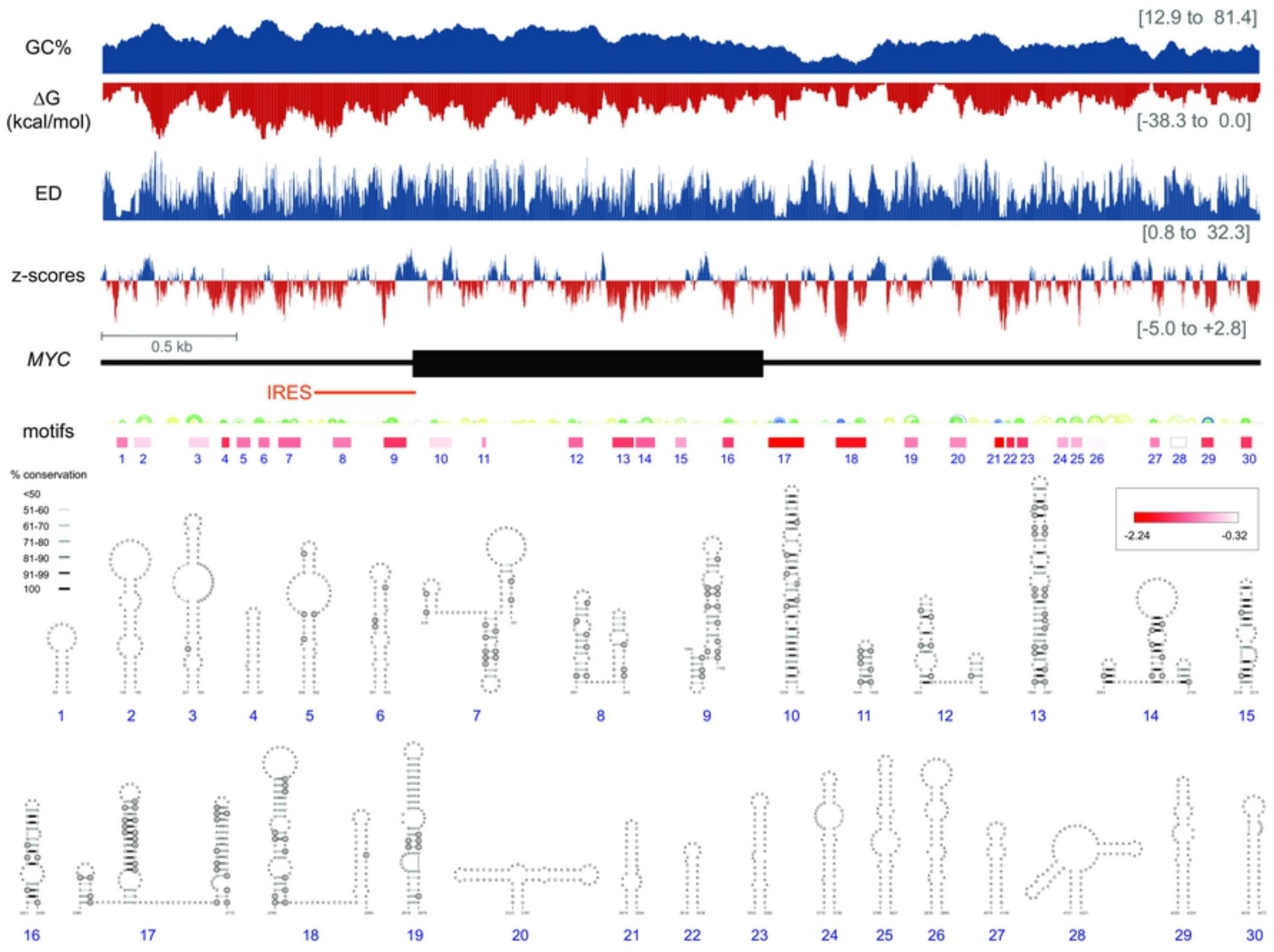


Figure 1