Decryption of sequence, structure, and functional features of SINE repeat elements in SINEUP non-coding RNA–mediated post-transcriptional gene regulation

Harshita Sharma[1], Matthew N Z Valentine[1], Naoko Toki[1,], Hiromi Sueki Nishiyori[1], Stefano Gustincich[2], Hazuki Takahashi[1]* and Piero Carninci[1,3]*

1 Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, 230-0045, Japan
2 Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genova, Italy
3 Human Technopole, Milan, 20157, Italy
* Corresponding authors
E-mails: carninci@riken.jp; piero.carninci@fht.org; hazuki.takahashi@riken.jp

Abstract

RNA structure folding largely influences RNA regulation by providing flexibility and functional diversity. In silico and in vitro analyses are limited in their ability to capture the intricate relationships between dynamic RNA structure and RNA functional diversity present in the cell. Here, we investigate sequence, structure and functional features of mouse and human SINE-transcribed retrotransposons embedded in SINEUPs long non-coding RNAs, which positively regulate target gene expression post-transcriptionally. In-cell secondary structure probing reveals that functional SINEs-derived RNAs contain conserved short structure motifs essential for SINEUP-induced translation enhancement. We show that SINE RNA structure dynamically changes between the nucleus and cytoplasm and is associated with compartment-specific binding to RBP and related functions. Moreover, RNA–RNA interaction analysis shows that the SINE-derived RNAs interact directly with ribosomal RNAs, suggesting a mechanism of translation regulation. We further predict the architecture of 18 SINE RNAs in three dimensions guided by experimental secondary structure data. Overall, we demonstrate that the conservation of short key features involved in interactions with RBPs and ribosomal RNA drives the convergent function of evolutionarily distant SINE-transcribed RNAs.

Introduction

In long non-coding RNAs (lncRNAs), evolutionary selection takes place not only at the level of primary sequence and expression control from syntenic loci, but also at the level of structure and function[1]. This leads to conservation of function despite divergent sequence and secondary structure among lncRNA orthologs[1] and highlights the need to look beyond the sequence to understand the relationship between sequence, structure, and function in lncRNAs. The presence of modular secondary structures in several functional lncRNAs[2-4] also supports the notion that lncRNA evolution favors the conservation of functionally relevant short stretches of sequence and/or structure that form distinct interacting domains for RNA–DNA, RNA–RNA, and RNA–protein regulatory complexes[5,6]. The combination of these different

38 domains provides the lncRNAs with their unique functional specificity and diversity[5]. Such "modular RNA
39 code" was validated in recent studies where short structure motifs were found to control the lncRNA
40 stability and function, such as 3' end triple helix in MALAT1 and NEAT1[7,8], 5' asymmetric G-rich internal
41 loop (AGIL) motif in mouse Braveheart (Bvht)[4], pseudoknots in human MEG3[9], and a terminal hairpin
42 stem-loop (SL1) in mouse lncRNA antisense to *Uchl1* mRNA[10-12]. In the antisense to *Uchl1*, the crucial SL1
43 region is within an embedded short interspersed nuclear element (SINE) from the B2 family, which forms
44 the effector domain (ED) that is essential for targeted mRNA translation upregulation[10,11].

45 SINEs are on average 100–300 bp in length and are non-autonomous retrotransposons that are
46 evolutionarily derived from RNA polymerase III transcripts, mainly 5S ribosomal RNA (rRNA), 7SL RNA, and
47 transfer RNA (tRNA)[13,14]. These elements, previously considered "junk DNA", are abundant in gene-rich
48 regions and their role in gene regulatory networks has been established[15-18]. Many SINEs are transcribed
49 as enhancer RNAs and form the regulatory domains of functional lncRNAs[19-22] as well as participating in
50 the nuclear retention of lncRNAs[23], STAU1-mediated mRNA decay[24], and promotion of sense mRNA
51 translation by antisense lncRNA[10], thus supporting the Repeat Insertion Domains of LncRNAs (RIDL)
52 hypothesis[6]. Yet the majority of SINE regulatory functions and structural features remain poorly
53 understood. In mouse genome, the B2 family of SINE constitutes approximately 2.4% of the genome with
54 presence of ~3.5 x 10^5 copies[25]. The SINEB2 family is broadly classified into B3A, B3, B2_Mm2, B2_mm1t,
55 and B2_mm1a sub-families in descending order of evolutionary age according to the Repbase repeat
56 elements database[26-28]. Synthetic antisense long non-coding RNAs SINEUPs which comprise a SINEB2
57 retrotransposon positively regulate the translation of the sense mRNA that they overlap[29,30]. Not only
58 several mouse SINEB2s but also human free right Alu monomer (FRAM) repeat element and human MIRb
59 are found in natural antisense lncRNAs and display post-transcriptional SINEUP-induced gene
60 regulation[10,31], despite their lack of similarity in terms of repeat length, sequence composition, and
61 predicted secondary structure.

62 Here, to comprehend more clearly the extent of the SINEUP family, we used a synthetic SINEUP system
63 to functionally screen 15 SINEB2 elements originally embedded in mouse natural antisense transcripts.
64 The proven importance of SINE repeats in functional lncRNA domains prompted us to study the primary
65 sequence and secondary structure of these SINE-derived RNAs. Although we have previously determined
66 the folding of antisense *Uchl1* SINEB2 in vitro from its chemical footprinting and NMR data[11,12], structural
67 studies in solution cannot fully grasp the structural dynamics and interaction with cellular machinery.
68 Moreover, to describe the relationship between sequence, structure, and function in SINE-derived RNAs,

69 we need to study the RNA structure of multiple SINE-derived RNAs in the cellular environment and its
70 specific compartments, because RNA can acquire different conformations depending on subcellular
71 location, functional state, and biomolecular interactions[32] . Additionally, we need to understand RNA
72 modifications in functional SINE-derived RNAs, specifically pseudouridylation, which is reported to be
73 crucial for the structural stability, flexibility, and function of RNAs involved in translation[33-35].

74 LncRNA structures are known to act as a molecular scaffold for proteins[36,37], and in fact, the binding of
75 heterogeneous nuclear ribonucleoprotein K (HNRNPK) with SINEUP and target mRNA complex is vital for
76 SINEUP-mediated upregulation of target mRNA translation[38]. Previously, we found that upon HNRNPK
77 knock-down, the co-localization of endogenous *Uchl1* mRNA and SINEUP RNA in the cytoplasm was
78 decreased which resulted in loss of translation upregulation by SINEUP-UCHL1[38]. This negative effect of
79 low HNRNPK expression on SINEUP function was also observed in SINEUP-GFP[38]. Furthermore, RNA
80 immunoprecipitation analysis showed that HNRNPK binds with SINEUP-EGFP mRNA complex in both
81 nucleus and cytoplasm and co-sediments with SINEUP and EGFP mRNA in free 40S, monosome and light
82 polysome fractions in polysome profiling[38]. In other words, HNRNPK not only aids in the export of SINEUP
83 and target mRNA complexes to the cytoplasm but also participates in SINEUP RNA-translation complex[38],
84 thus, exhibiting two discrete functions in the nucleus and the cytoplasm. This knowledge makes it
85 imperative to examine how specific SINE RNA structures in different cellular compartments interact with
86 RNA binding proteins (RBPs) and support their distinct functions. In addition to intramolecular structure,
87 intermolecular interactions with other RNAs play a critical role in gene regulation by non-coding RNA[40,41],
88 notably the binding to rRNAs by linear and circular internal ribosome entry sites (IRESs) for promotion of
89 translation[42,43]. In fact, our previous molecular dynamics simulation study on NMR solution structure of
90 SINEB2 SL1 identified motifs in three-dimensions (3D) similar to those in rRNAs[44]. This data and the
91 association of SINEUPs with ribosomal subunits in polysome profiles[38] made us ponder the possibility that
92 direct interaction between SINE-derived RNAs and rRNAs may be a mechanism by which SINEUPs enhance
93 mRNA translation.

94 To solve the puzzle of SINE-transcribed RNA structure, we utilized a method called in vivo click selective
95 2'-hydroxyl acylation and profiling experiment (icSHAPE), which chemically probes flexible RNA bases
96 inside living cells and thus provides an advantage over other chemical or enzymatic methods[45]. We
97 present two-dimensional (2D) secondary structure models for 17 mouse SINEB2s-derived RNAs and a
98 human SINE FRAM RNA, providing for the first time experimentally verified in-cell 2D-structure models of
99 evolutionarily close and distant SINE-transcribed RNAs at this scale. We found that structural deletion

100 leading to the gain of an apical stem-loop (known as SL1) restores the activity in a SINEUP that is
101 functionally dormant (i.e., lacking the ability to significantly enhance the target mRNA translation). Based
102 on the 2D models, we discovered seven short conserved-structure motifs among the SINE RNAs tested,
103 three of which harbor modified RNA bases. The icSHAPE-guided structures for the nucleus and cytoplasm
104 showed that SINE RNAs in both mouse (SINEB2, derived from tRNA) and human (FRAM, derived from 7SL
105 RNA) rearrange into a cloverleaf-like structure in the cytoplasm and have partially conserved structure
106 motifs. We also observed that compartment-specific structures of SINE RNA are associated with
107 distinctive binding patterns of HNRNPK in the nucleus and the cytoplasm by using a single-end enhanced
108 crosslinking and immunoprecipitation assay (seCLIP)[46]. Additionally, when motifs were compared in silico
109 with those in the Protein Data Bank (PDB), we detected a remarkable similarity of SINEUP 2D structure
110 motifs with rRNA structure motifs. We also checked RNA–RNA pairs by the psoralen analysis of RNA
111 interactions and structures (PARIS2) method[41] and identified interactions of mouse and human SINE-
112 derived RNAs with human 18S and 28S rRNAs in the conserved structure motif regions, which shed some
113 light on SINEUP's mechanism of action. Finally, we used our experimental 2D structure data to guide the
114 prediction of structure models for 18 SINE-transcribed RNAs in 3D which can provide a base for future
115 experimental 3D structure studies. In brief, this study demonstrates that despite sequence variability
116 among SINE-transcribed RNAs, short stretches of conserved sequence and structure preserve the
117 regulatory functions by providing binding sites for RBPs and rRNA. These short motifs are likely to be
118 stabilized by modifications and non-canonical base interactions and are evolutionarily favored to maintain
119 the SINEUP functionality even in distantly related SINE-derived RNAs.

120

121 ## Results

122 ### The SINEUP effect is a prevalent functional feature of mouse SINEB2s-derived RNAs

123 To experimentally verify the translation-regulatory function of various mouse SINE-transcribed RNAs in
124 the SINEUP effector domain (ED), we used a previously established synthetic SINEUP-GFP model that
125 enhances EGFP mRNA translation (**Fig. 1a**). We swapped the SINEB2 from antisense *Uchl1* in SINEUP-GFP
126 with those from other mouse antisense to genes: *Txnip*, *Gadd45α*, and *Uxt* (**Supplementary Table 1a**).
127 RepeatMasker revealed the presence of an inverted SINEB2 repeat spanning from 1389 to 1495
128 nucleotide positions (107 bp long) in antisense *Txnip*, and both antisense *Gadd45α* and antisense *Uxt*
129 lncRNAs contain two inverted SINEB2 repeats in the region not complementary to the sense mRNA. Both

130 antisense *Uxt* SINEB2 sequences are from the B3 sub-family; in the FANTOM3 cDNA clone, one sequence
131 covered positions 160 to 290 (131 bp) and the other 774 to 960 (187 bp) (**Supplementary Table 1a**). Based
132 on their natural order of occurrence from the 5′ end of the transcript, the sequences were named
133 antisense *Uxt* SINEB2-a (131 bp, proximal to the 5′ end) and antisense *Uxt* SINEB2-b (187 bp, distal to the
134 5′ end) (**Supplementary Table 1a**, **Fig. 1b**). In contrast, both SINEB2 repeats of antisense *Gadd45α* belong
135 to the B2_Mm2 sub-family (according to RepeatMasker nomenclature). We named these two sequences
136 antisense *Gadd45α* SINEB2-a (131 bp, proximal to the 5′ end) and antisense *Gadd45α* SINEB2-b (114 bp,
137 distal to the 5′ end) (**Supplementary Table 1a**). To understand whether both SINEB2 sequences are
138 essential for SINEUP function and how their order affects SINEUP activity, we designed antisense *Gadd45α*
139 and antisense *Uxt* constructs with a or b sequences and also with both sequences in the natural (ab) or
140 the inverted (ba) order (**Fig. 1b**). The shortest SINEB2 of those tested, antisense *Txnip* SINEB2 (107 bp),
141 enhanced the GFP protein level (**Fig. 1c** and **d**, **Supplementary Table 1**). None of the four constructs with
142 antisense *Gadd45α* SINEB2s could exert the SINEUP effect, regardless of the order of SINEB2 a and b (**Fig.**
143 **1c** and **d**). Indeed, natural antisense *Gadd45α* was found to inhibit translation of sense *Gadd45α*
144 (**Extended Data Fig. 1**), which implies that the SINEUP activity of antisense *Gadd45α* SINEB2 RNA is
145 dormant, possibly due to the lack of functional regions. Interestingly, of the antisense *Uxt* SINEB2 repeats,
146 only SINEB2-b upregulated the GFP protein level significantly, and when the position of the two SINEB2
147 sequences was switched from the natural order (ab switched to ba), their SINEUP effect was obliterated
148 (**Fig. 1c** and **d**). As negative controls, in addition to empty vector, we tested direct antisense *Uchl1* SINEB2,
149 where the orientation of SINEB2 was flipped from the one in the functionally active construct, and the
150 deletion mutant of antisense *Uchl1* SINEB2 (ΔAS *Uchl1* SINEB2) in which the deletion of SINEB2 shifts
151 downstream synthetic and non-SINEB2 annotated sequences in its place in the ED of SINEUP[10]. Consistent
152 with our previous work[10], none of these negative controls could upregulate GFP translation which
153 demonstrates that SINE RNA region is essential for SINEUP function (Fig. **1c** and **d**). Upholding the post-
154 transcriptional regulatory nature of SINEUPs, none of the SINEUPs changed the EGFP mRNA expression
155 level significantly (**Extended Data Fig. 2a**).

156 To further expand analysis of the SINEB2-derived RNAs present in mouse natural antisense transcripts
157 (**Supplementary Table 1**), we used the shortest functional model of SINEUP-GFP, named miniSINEUP-
158 GFP[29]. SINEB2 sequences from antisense *Uchl1* and antisense *Uxt* enhanced GFP protein expression by
159 around 1.8–2.2-fold (**Fig. 1e**, Lanes 2 and 3 in the left graph). We additionally found that combining the
160 SINEB2 from antisense *Uchl1* with antisense *Uxt* SINEB2-b upregulated GFP protein production; however,
161 the protein fold change was lower than the effect of either SINEB2 independently (**Fig. 1e**, Lane 4 in the

162 left graph). We additionally tested three repeats of antisense *Uchl1* SINEB2 but the SINEUP function
163 remained similar to that with only one repeat, implying that these repeats do not function in a synergistic
164 or additive manner (**Fig. 1e**, Lane 5 in the left graph).

165 Since the SINEB2 sub-family B3 acted as a functionally effective SINEUP ED, we decided to check the
166 SINEUP potential of consensus subfamily B3 sequence from the Repbase database[27,47]. Indeed, the
167 sequence is functionally active against GFP (~1.6-fold change) (**Fig. 1e**, Lane 6 in the left graph), but not
168 to the level of antisense *Uchl1* SINEB2 RNA, which is a similar result to tests on other SINEB2 sequences
169 in the B3 sub-family from antisense lncRNAs to genes: *Nars2*, *Abhd11*, *Epb4.9*, *Wfdc5*, *Pgbd1*, and *Gsk3b*
170 (**Fig. 1e**, Lanes 7–10 in the left graph and Lanes 3–4 in the right graph).

171 The failure of SINEUPs with antisense *Gadd45α* SINEB2 from the B2_Mm2 sub-family prompted us to test
172 whether the SINEUP effect is limited to the B3 sub-family only. To confirm this, we investigated two
173 B2_Mm2 sub-family sequences (from antisense *Rhod* and antisense *E4f1*), one B2_Mm1t (from antisense
174 *E4f1*) and one B3A (from antisense *Rhod*) (**Fig. 1e**, Lanes 5–8 in the right graph) and found that all the
175 SINEB2 sequences, regardless of their sub-family type, length, or parent natural antisense transcripts,
176 upregulated GFP protein expression to more or less the same extent, which reveals the vastness of the
177 SINEUP family. Similar to that of long SINEUPs, this miniSINEUP effect was also manifested at the post-
178 transcription level (**Extended Data Fig. 2b**). Although the SINEUP RNA level for SINEB2s from antisense
179 *Wfdc5*, *Pgbd1*, and *Gsk3b* was unusually high (3–10-fold higher than other miniSINEUPs), the GFP mRNA
180 expression level in all miniSINEUPs remained unaltered.

181

182 **Functional SINEs form double-stranded structures in the cell and consist of conserved short structure**
183 **motifs**

184 To check any common sequence features between different SINEB2-transcribed RNAs showing the
185 SINEUP effect and to identify any functionally important regions, we performed multiple sequence
186 alignment using ClustalW (**Extended Data Fig. 3a** and **b**). This revealed conserved regions of RNA pol III
187 internal promoters known as A box and B box[48] (**Extended Data Fig. 3a**) along with sequence diversity
188 between SINEB2 elements tested in this study. Moreover, in the phylogenetic tree based on the multiple
189 sequence alignment, the functionally active SINEB2s that significantly upregulated GFP translation were
190 not clustered separately from the inactive SINEB2s that failed to do so (**Extended Data Fig. 3b**). In other
191 words, on the basis of sequence, the functional SINEB2 RNAs do not possess any unique features that

192 confer a functional advantage over the inactive SINEB2 RNAs. This lack of conserved sequence domains
193 led us to examine the in-cell secondary structure of SINEB2s by using the icSHAPE method to map the RNA
194 secondary structure inside living cells[45,49].

195 We compared the icSHAPE antisense *Uchl1* SINEB2 RNA secondary structure with our previous in vitro
196 chemically probed structure[11], which had identified four internal-loops (IL1–IL4) and three stem-loops
197 (SL1–SL3) (**Fig. 2a**). Both methods gave a notably similar double-stranded structure for antisense *Uchl1*
198 SINEB2 RNA (**Fig. 2a**). The icSHAPE-based structure comprises three internal loops similar to the IL2, IL3,
199 and IL4 found in the chemical footprinting-based structure, with slight differences in the sequence
200 composition. The in-cell structure lacks SL3, which indicates the influence of intracellular factors and
201 dynamic structural changes compared to the less complex conditions in vitro. Interestingly, SL1 and SL2
202 are conserved between the structures derived by these two different methods, which suggests that they
203 represent the most dominant and stable structure components in the ensemble. The structural
204 conservation is also understandable knowing that SL1 is crucial for SINEUP function[11]. In a similar way,
205 secondary structure modelling was performed for other SINEB2s-derived RNAs (**Fig. 2b** and **c**, **Extended**
206 **Data Fig. 4**) and revealed that they all retained double-stranded features with a terminal GC-rich stem-
207 loop similar to SL1 of antisense *Uchl1* SINEB2, even though all SINEB2 RNAs tested here had diverse global
208 structures with variable sequence length and composition.

209 The importance of short structure motifs like SL1 in the SINEUP function prompted us to look for other
210 local motifs that are shared between different SINEB2-transcribed RNAs. For this purpose, we utilized the
211 ExpaRNA tool[50] and, using the antisense *Uchl1* SINEB2 RNA structure as a base, searched for its exact
212 sequence and structure patterns in other icSHAPE-derived SINEB2 RNA structures. From this search, we
213 discovered seven GC-rich structure motifs partially conserved among the SINEB2 RNA structures tested in
214 this study (**Fig. 2a**–**c**, **Extended Data Fig. 4**). Motifs 2 and 3 are comprised of 'AGG' trinucleotide(s), while
215 motifs 4, 5, 6, and 7 are 'UGG' enriched. Motif 6 lies within the SL1 region making it a potentially functional
216 motif. Furthermore, we wondered if the motifs are conserved in SINE-transcribed RNAs that are
217 evolutionarily distant from mouse SINEB2 elements, such as human SINE FRAM, which has been proven
218 to be effective as a SINEUP against both endogenous and exogenous mRNA targets[31]. Indeed, the
219 structure motifs in the whole-cell icSHAPE structure of FRAM RNA provided five matching regions (**Fig.**
220 **2d**). This hints that conservation of such short motifs may contribute in functional preservation. We
221 further examined the PDB database for 3D RNA structure motifs that were akin to our SINEUP structure
222 motifs. For this purpose, we employed a sequence- and structure-specific search for each of the seven

223 SINEUP motifs in the RNA Characterization of Secondary Structure Motifs (CoSSMos) database[51]. This
224 resulted in hits that matched 16S rRNA (PDB ID: 5JB3) and 18S rRNA (PDB ID: 5A2Q) to 2D structures for
225 SINEUP motifs 2 and 6, respectively (**Extended Data Fig. 5a**). Interestingly, the matching motifs are very
226 similar to in-cell derived SINEUP motifs in terms of sequence and structure. On closer inspection, we
227 observed that the matching region to 18S rRNA is within helix 34, which forms the head of the mRNA
228 entry channel[52] (**Extended Data Fig. 5b**).

229

230 **Acquisition of SL1-like structure turns a dormant SINEUP active**

231 Considering the conservation of short local structure motifs in functional SINE-transcribed RNAs, we
232 theorized that dormant SINEUPs, here the SINEB2s that failed to significantly upregulate GFP protein
233 expression in **Fig. 1c**, might include embedded inhibitory regions that prevent their proper folding into
234 functional domains. To test this, we chose a functionally inactive SINEB2 sequence, antisense *GADD45α*
235 SINEB2-a, and checked the effect of deletion of different sequence regions on its structure. In total, we
236 created four deletion mutants at positions 27–36, 40–50, 66–76, and 92–101 from the 5' end of SINEB2
237 including two tRNA-derived conserved internal RNA polymerase III promoter regions, A-box and B-box
238 (**Extended Data Fig. 3a**). We observed structural changes caused by these deletions by in-silico RNA
239 structure prediction which we later confirmed in the cell by icSHAPE. Analysis of in-cell structures of
240 deletion mutants in comparison to that of wild-type antisense *GADD45α* SINEB2-a and antisense *Uchl1*
241 SINEB2 (**Fig. 2a**, **e**, **Extended Data Fig. 6**) showed that out of all the mutants, Δ27–36 shares the most
242 sequence and structural similarity in SL1 structure and motif 6 with functional antisense *Uchl1* SINEB2
243 (**Fig. 2a**, **e**). The wild-type antisense *Gadd45α* SINEB2-a carries a deformed SL1 where the stem is
244 interrupted by multiple internal loops (**Fig. 2e**). Upon deletion of 27–36 nucleotide positions, this stem
245 becomes less structured and transforms into a properly folded SL1 similar to antisense *Uchl1* SINEB2. The
246 Δ40–50 mutant, though shares similarity in overall 2D structure in terms of position of motifs 1, 2, 4, 5,
247 and 7 with Δ27–36 mutant, lacks a proper SL1 and no motif 6 (**Extended Data Fig. 6**. On the other hand,
248 Δ66–76 and Δ92–101 mutants contain partial motif 6 but have truncated SL1 (**Extended Data Fig. 6**). This
249 disparity in structure is reflected in their SINEUP function. When we co-transfected the mutants with
250 sense EGFP into HEK293T/17 cells, as done in the previous SINEUP-GFP functional analysis experiment we
251 observed that out of the five mutants Δ27–36 could significantly upregulate GFP protein level compared
252 to full-length antisense *GADD45α* SINEB2-a (**Fig. 2f**). Deletion mutant Δ40–50 and Δ66–76 could not
253 upregulate GFP translation while Δ92–101 mutant in some experiments showed around 1.5-fold increase

254 in GFP protein but had inconsistent performance in others (**Fig. 2f**). This suggested that some structural
255 features in the wild-type antisense *GADD45α* SINEB2-a are detrimental to function, whereas the
256 acquisition of short local structure motifs leads to functionally favorable RNA folding, which can restore
257 SINEUP function. Especially, properly folded SL1 and motif 6 are essential not only for exerting the SINEUP
258 effect but also for ensuring functional stability.

259

260 **SINEUP structure dynamics across nucleus and cytoplasm**

261 Our SINEUP icSHAPE data from whole cell lysate showed a set of common motifs among different SINEB2s,
262 yet it was an average of whole-cell structure ensemble and did not help to infer structure dynamics within
263 the cell. RNA structure is closely related to cellular processes and can acquire distinct conformations
264 depending on the cellular compartment[32]. Understanding the different RNA structures becomes more
265 crucial in the case of SINEUPs because they are involved in post-transcriptional translation regulation, a
266 function specific to cytoplasm. Therefore, it is logical to examine the structure dynamics of functional
267 SINEB2s-derived RNAs both in the nucleus and the cytoplasm, to better understand the contribution of
268 structural conformations to the biological processes specific to these subcellular regions. To achieve this,
269 we made icSHAPE libraries from nucleoplasmic and cytoplasmic fractions of cells transfected with SINEUP-
270 GFP carrying either mouse antisense *Uchl1* SINEB2 or human SINE RNA FRAM. The 2D structure models
271 based on this icSHAPE data revealed that the base of the SINEB2 RNA structure is mostly stable between
272 the nucleus and the cytoplasm, with only a slight variation observed around the region of motif 2, being
273 slightly more open in the nucleus than in the cytoplasm (**Fig. 3a**, **b**), and a similar structure difference in
274 human FRAM-transcribed nuclear RNA (**Fig. 3c**, **d**). In addition, the internal loop between motifs 5, 6, and
275 7 in the nuclear mouse SINEB2 RNA goes through a structural rearrangement in the cytoplasm and
276 transforms into a cloverleaf-like structure (**Fig. 3a**, **b**). In the case of human FRAM RNA, both nuclear and
277 cytoplasmic structures are quite similar and form a cloverleaf-like structure, though the cytoplasmic
278 structure is slightly more structured (**Fig. 3c**, **d**). This could be possibly due to the transient nature of any
279 alternative structures, as icSHAPE structure represent only the average of the structure ensemble.

280 The nucleus-specific internal loop structure might play a role in the export of SINEUP RNA to the
281 cytoplasm, whereas the cytoplasmic cloverleaf-like structure with SL1 might be crucial for interaction with
282 translation machinery. Interaction with HNRNPK is crucial in both of these SINEUP processes[38]. To check
283 how the different nuclear and cytoplasmic structures of SINEB2 RNA are associated with HNRNPK, we

284 prepared seCLIP[46] libraries from the nuclear and cytoplasmic fractions of SINEUP-GFP co-transfected cells
285 (**Fig. 3e**). The results showed that in the nucleus HNRNPK interacts with short stretches of SINEB2
286 coinciding in motifs 1, 3, and 4 (**Fig. 3b**), but in the cytoplasm HNRNPK occupied a long double-stranded
287 region covering the 5' end of the stem in all the motifs except motif 7. Such disparate interactions could
288 possibly arise from distinct functions of HNRNPK in SINEUP RNA transport and in translation that are
289 specific to the cellular compartment.

290

291 **SINEUP structure motifs interact with rRNA**

292 We wanted to further investigate the role of SINEB2 RNA structure in translation because it has features
293 reminiscent of tRNA; it forms a cloverleaf-like structure in the cytoplasm and possesses two pseudouridine
294 (Ψ) sites at 16 bp and 112 bp (**Extended Data Fig. 7**). CircRNA IRESs[43] and many other IRESs are known to
295 interact with 18S rRNA[42,53] to induce cap-independent translation, however the region by which SINEUP-
296 RNA interacts with ribosomal subunits and polysomes is not known.

297 Considering the similarities in sequence and structure, we wondered whether such motifs in SINEB2 RNA
298 can interact with rRNA. To get a better understanding of SINEUP–rRNA interactions, we employed PARIS2,
299 which has proved useful in capturing direct mRNA–rRNA short interactions[41,54]. We prepared PARIS2
300 libraries from HEK293T/17 cells transiently expressing SINEUP-GFP (antisense *Uchl1* SINEB2) or
301 miniSINEUP-GFP (FRAM) along with their target sense EGFP mRNA and sequenced them on a next-
302 generation sequencing platform. We could capture few incidents of SINEUP-rRNA pairs (**Fig. 4**, **Fig. 5**,
303 **Extended Data Fig. 8**), which is understandable given the technical limitation of psoralen crosslinking and
304 inefficient detection of scarcely expressed RNA duplexes and highly dynamic interactions in PARIS2[41,55].
305 Nevertheless, it was intriguing to see interactions of mouse SINEB2 element with human rRNA in both
306 small and large subunits of ribosome (**Fig. 4a**, **b**, and **c**). We discovered that the 3' side of motifs 1 and 2
307 of SINEB2 RNA interacted with a region in helices 14 and 15 of 18S rRNA (Int 1 in **Fig. 4a** and **c**). On the
308 other hand, the region between motifs 5 and 6 formed contacts with domain III helices 55, 56, and 59 of
309 28S rRNA. Motif 5 and the flanking region also interacted with domain I, helix 25, expansion segments
310 ES7c and ES7d of 28S rRNA (Int 2 and 3 in **Fig. 4b** and **c**).

311 Interestingly, functionally active human SINE-derived RNA FRAM displayed conservation of this rRNA
312 interaction trait. With better expression of FRAM RNA, we were able to detect more instances of FRAM-
313 rRNA interaction than for SINEB2 RNA. Similar to SINEB2, the 3' side of motifs 1 and 2 and of motifs 5 and

314 6 of FRAM RNA formed contacts with helix 15 of 18S rRNA (Int 2 in **Fig. 5a**, **b**, **c**). In addition, the FRAM
315 RNA stem loop analogous to SL1 (containing motifs 2/3, 5, and 6) interacted with 18S rRNA helix 21, es6b
316 and es6c (Int 1 in **Fig. 5a**, **b**, **c**), and helices 37 and 38 paired with the 3' end region of FRAM RNA (Int 3 in
317 **Fig. 5a**, **b**, **c**).

318 Moreover, the 5' side of FRAM RNA flanking motif 1 interacted with 28S rRNA domain III helix 48 and 49a
319 (Int1 in **Fig. 5d**, **e**, **f**), and helix 50 and 52 (Int5 **Fig. 5d**, **e**, **f**), whereas FRAM SL1 region interacted with 28S
320 rRNA domain IV, helix 63, ES27b and with domain I, helix 25, ES7c (Int 2 and 3 in **Fig. 5d**, **e**, **f**). The short
321 stem-loop at the 3' end of FRAM structure was involved in interaction with 28S rRNA domain V, helix 81
322 (Int4 in **Fig. 5d**, **e**, **f**).

323 The presence of SINEUP-rRNA pairs in PARIS2 data in addition to our knowledge of SINEUPs' fate in the
324 polysomes[38], that over 85% of the cytoplasmic SINEUP-GFP co-sediments with EGFP mRNA and HNRNPK
325 in free/40S ribosomal subunit and monosome fractions, supports the notion of SINEUP-rRNA
326 interaction[38].

327

328 **SINE RNA 3D structure models reveal non-canonical intramolecular interactions**

329 The SINE 2D structure models helped us to understand the conservation and functional importance of
330 local short motifs in maintaining the SINEUP effect by interacting with RBP and rRNA. However, they
331 lacked information about non-canonical intramolecular interactions that stabilize the RNA structure and
332 affect the availability of interaction sites in 3D. This results in canonically unpaired regions in the 2D
333 structures that are actually non-canonically paired in 3D[56]. Therefore, we took advantage of in vivo SINE
334 RNA 2D structure information and incorporated it into the machine translation of 3D structure models.
335 Thus, we created 3D models for 17 mouse and one human SINE-derived RNAs (**Fig. 6** and **Extended Data**
336 **Fig. 9**). The 3D models revealed that the SINEUP motifs are stabilized by base-stacking interactions along
337 the double-helical structure and that bases in the internal loop and stem-loop regions form non-canonical
338 interactions adding more complexity to the structure (**Fig. 6 a**,**c**,**e**,**g** and **Supplementary Tables 2** and **3**).
339 The molecular surface view of the nuclear SINEB2 RNA shows an elongated helical structure where SL1
340 takes the apical position (**Fig. 6b**). Interestingly, human nuclear FRAM RNA, which has a cloverleaf-like
341 conformation in 2D (**Fig. 3d**), folds similarly to nuclear SINEB2 RNA in 3D, with the three arms positioned
342 in such a way that all the motifs are in the central arm, SL1 is at the apex, and the other two arms are
343 close to the central one (**Fig. 2d** and **Fig. 6 e**,**f**). In the cytoplasm, both SINEB2 and FRAM RNAs fold into a

344 cloverleaf-like conformation where SL1 moves from the center to the left position and in FRAM RNA the
345 other two arms open up (**Fig. 6 c**,**d**,**g**,**h**). We also analyzed the 3D conformation of functionally inactive
346 antisense *Gadd45α* SINEB2-a and its deletion mutants in comparison to antisense *Uchl1* SINEB2 (**Extended**
347 **Data Fig. 10**). The 3D models showed that among all the mutants, the functional Δ27–36 mutant adopts
348 a similar global structure as well as SL1 and motif 6 sub-structures with that of antisense *Uchl1* SINEB2 in
349 3D space (**Extended Data Fig. 10 a,c**) consistent with the 2D structure data. The wild-type antisense
350 *Gadd45α* SINEB2-a and functionally inactive mutants Δ40–50, Δ66–76 and Δ92–101 which have deformed
351 SL1, vary in RNA 3D architecture from the functional mutant (**Extended Data Fig. 10 a,b,d**–**f**).

352

353 **Discussion**

354 There are about 175 families of SINE scattered across different domains of life[57]. Although generally
355 considered as parasite elements of the genome, a growing number of studies point to their gene-
356 regulatory activity, such as enhancer-like function, polyadenylation signal, transcription factor binding
357 sites, mRNA nuclear retention, and translation upregulation[10,15-18,23]. However, there are no extensive
358 studies on the sequence and structural features of functional SINE-transcribed RNA elements. Here, we
359 examined the SINEUP activity of 15 as-yet-uncharacterized SINEB2-derived RNAs from mouse natural
360 antisense transcripts and assayed their function in GFP protein upregulation. Most (12 out of 15) of the
361 SINEB2 RNAs showed SINEUP activity, but the remaining 3 (antisense *Gadd45α*-SINEB2-a and -b, and
362 antisense *Uxt* SINEB2-a) failed to upregulate GFP protein expression significantly under the current
363 experimental settings. The shortest SINEB2 RNA to enhance GFP protein expression, antisense *Txnip*
364 SINEB2, was found to be 107 nucleotides long, which is half the length of the longest functional natural
365 SINEB2-derived RNAs tested here (antisense *Nars2* SINEB2 and antisense *Epb4.9* SINEB2). This suggests
366 that the SINEUP class of RNAs in nature is vast and diverse and is likely to be a much larger class of
367 antisense RNAs than yet explored. Functional observation of the consensus SINEB2 sub-family B3
368 sequence also supported this, although it enhanced GFP translation to a lesser extent than the highly
369 functional antisense *Uchl1* SINEB2 RNA. This might be because consensus sequences represent the most
370 frequent occurrence of a particular nucleotide in a multiple sequence alignment of related sequences,
371 which may not necessarily be functionally optimal. We observed that in most cases only one repeat of
372 SINEB2 RNA as the effector domain is sufficient to exert the SINEUP effect, possibly because two SINEB2-
373 derived RNAs (which are generally GC-rich) may share some short complementary regions that can
374 interact and hinder the proper RNA folding into key functional features or may compete for binding

375 proteins and RNA partners. Another interpretation of this non-synergistic nature of these functional
376 SINEB2 RNAs is that SINEUP-induced protein upregulation is used to fine-tune translation in the cell,
377 keeping the protein within a physiological range of up to 1.5–3-fold change. As previously observed, this
378 fold change is sufficient to rescue haploinsufficient gene dosage and disease phenotype in vivo by
379 upregulation of protein expression of the subunit 7B of cytochrome c oxidase (*cox*7b) in a medaka fish
380 model of human microphthalmia with linear skin defects (MLS) syndrome[58], and rescue of endogenous
381 defective Frataxin (FXN) expression in a cellular model of Friedreich's ataxia[59] as well as rescue of
382 neurodegeneration and motor phenotype by enhancing endogenous Glial cell-derived neurotrophic
383 factor (GDNF) protein expression in a mice Parkinson's disease model[60]. Incidentally, this characteristic
384 makes SINEUPs ideal for therapeutic studies where proteins need to be regulated within a physiological
385 range. The sequence diversity within the same B3 sub-family or in comparison with other sub-families
386 such as B3A, B2-Mm2, and B2_Mm1t (as per Repbase nomenclature)[27] does not prevent the SINEUP
387 function. This observation confirms the universality of the SINEUP function in the SINEB2 repeat family
388 and the flexibility of its members to maintain key functional features despite varying length and
389 composition of RNA sequences. It would be interesting to explore other families of SINE elements in other
390 species and across a wide variety of mammalian and eukaryotic genomes in the future.

391 Noting the sequence diversity of functional SINEB2-derived RNAs and the importance of RNA structure in
392 non-coding RNA functions, we looked at the role of the SINE RNA secondary structure in the SINEUP
393 phenomena in living cells by using icSHAPE experiments. Consequently, we noticed a high correlation
394 between the in-cell 2D structure of antisense *Uchl1* SINEB2 RNA and the previously reported structure
395 from chemical footprinting[11]. Moreover, we found that the key functional SL1 structure was conserved
396 among various SINEB2s RNAs. When the functionally dormant antisense *Gadd45α* SINEB2-a RNA that lacks
397 an SL1 region was mutated to change its structure, one of the mutants (Δ27–36) acquired a properly
398 folded SL1 structure and a 2D structure similar to the icSHAPE structure of antisense *Uchl1* SINEB2 RNA,
399 resulting in SINEUP activity gain. This further emphasizes the functional significance of this short structure
400 in SINEUP-induced gene regulation.

401 We discovered 7 conserved structure motifs among the icSHAPE data-driven 2D structure models of all
402 the SINEB2-derived RNAs tested here, though only partial matches in sequence were observed, given the
403 sequence variation. Since we observed partial matches to these structure motifs in the functionally
404 inactive SINEs also, we hypothesize that the SINEUP function is influenced not just by the mere presence
405 of these structure motifs but also by their relative positions on the structure. Some structures might be

406 sterically more exposed to interaction with molecular machinery in the 3D space. Additional structural
407 regions might help in proper folding of these motifs and play an accessory role in maintaining the SINEUP
408 function. Thus, any deletion might inadvertently disrupt functional structures and interaction with RBPs.

409 We recently reported that deletion of certain regions in antisense *Uchl1* SINEB2 RNA is detrimental to
410 SINEUP function[12]. Several of these deleted regions correspond to the SINEUP structure motifs identified
411 in this study: Δ100-111 disrupts motif 7, Δ130-141 is the deletion of the 3' end side of motifs 3 and 4, and
412 Δ5-14 disrupts motifs 1 and 2[12]. Interestingly, their Δ65–74 mutant[12], which deletes part of motif 6, does
413 not show a significant decrease in the SINEUP function, whereas another study demonstrated that
414 deletion of SINEB2 RNA SL1 consisting of the entire motif 6 abolishes SINEUP activity[11]. The in-cell
415 structure of human SINE-derived FRAM RNA also showed a partial match for SINEUP motifs and the
416 presence of a structural feature analogous to SL1. Furthermore, we found similarity between the structure
417 of the motif 6 sequence in 2D and that of a motif in 18S rRNA in 3D, and the structure of motif 2 was found
418 to be similar to a region in 16S rRNA in 3D. The matching region in 18S rRNA belongs to helix 34, which
419 forms the roof of the mRNA entry channel. The similarity of the SL1 region of SINEB2 RNA with rRNA was
420 previously noticed in a molecular dynamics simulation based on NMR data[44], and here in-cell analysis
421 again showed that the conserved and functionally crucial SINEUP motifs were similar to rRNA structure
422 motifs.

423 Mouse SINEB2s originate from tRNA element[57,61], and pseudouridylation of tRNA bases is reported to be
424 crucial for its stability and function[35]. Not only tRNAs, but all the RNAs involved in translation contain
425 several Ψ-modified sites[34,35]. This was also found to be the case for SINEUPs in this study. We identified
426 two Ψ modifications at positions 16 and 112 of antisense *Uchl1* SINEB2 RNA. The Ψ at position 112 is
427 inside motif 5, and that at position 16 is just at the motif 2 boundary. This indicates that pseudouridylation
428 of these SINEB2 RNA positions might help to stabilize the structure of these motifs by forming extra
429 hydrogen bonds by N1-H position of Ψ or by base-stacking, which is an intrinsic property of Ψ[34,62].

430 We further analyzed the structural dynamics of both SINEB2 and FRAM RNAs in the nucleus and cytoplasm
431 and identified two regions that differ between their nuclear and cytoplasmic structures, with motif 2 and
432 the region between motifs 5 and 6 of SINEB2 RNA slightly more open in the nucleus than in the cytoplasm.
433 Less structured regions were also observed in the nuclear structure of FRAM RNA. We wondered whether
434 these nucleus-specific structures can provide binding sites for proteins involved in nuclear transport. We
435 knew that binding to HNRNPK is crucial for cytoplasmic co-localization of SINEUP-GFP and sense mRNA
436 complex and their association with polysomes[38]. Furthermore, HNRNPK is reported to interact with c-myc

437 IRES and stimulate its cap-independent translation[63]. Since HNRNPK participates in both nuclear transport
438 and translation regulation, we checked if structurally dynamic regions have any effect on its interaction in
439 subcellular compartments. Indeed, we observed distinctive binding patterns of HNRNPK in the nucleus
440 and cytoplasm; in the nucleus, it interacts with short regions in motifs 1, 3, and 4 of SINEB2 RNA, but in
441 the cytoplasm, it interacts with a much longer double-stranded region covering all but one motif. These
442 particular binding patterns could be connected to different functions of HNRNPK in the nucleus and
443 cytoplasm.

444 In the cytoplasm, the loop region between motifs 5, 6, and 7 of mouse SINEB2 RNA assumes a cloverleaf-
445 like structure, formed from a multiloop where four stems converge. One stem is SL1, which harbors motif
446 6, another stem consists of motif 7, and a disrupted motif 5 makes the top of the central stem where the
447 Ψ at position 112 is situated at the multiloop branching point. Interestingly, though the nuclear and
448 cytoplasmic structures of FRAM RNA are highly similar, we discerned a similar cloverleaf-like structure.
449 icSHAPE structures are generally enriched for the most stable and dominant structural conformations in
450 the cell, so any alternative short-lived structures might not have been captured. It is fascinating to see
451 that mouse SINEB2 RNA that originated from tRNA and human FRAM RNA that originated from 7SL RNA[61]
452 share common characteristics: short conserved structural motifs, a cloverleaf-like structure in the
453 cytoplasm, and positive regulation of translation. Previously, we created a minimal functional SINEUP by
454 deleting nucleotide positions 1-30 and 120-167 from SINEB2 RNA, thus keeping only nucleotide positions
455 31-119 which is the region corresponding to motifs 5, 6, and 7 as well as SL1 and cytoplasmic cloverleaf-
456 like structure[12]. This mutant upregulated GFP mRNA translation and retained about 80% of full length
457 SINEB2 functional activity supporting the significance of this structural region in SINEUP function[12].

458 In addition to the intramolecular structure, we were interested to better understand the intermolecular
459 interactions of SINE-derived RNAs with translation machinery. Interaction with 18S rRNA has been
460 reported to be crucial for both linear and circular IRESs[42,43,53]. Knowing that SINEUP-GFP RNA co-
461 sediments with free small subunits of ribosomes and monosomes, we checked direct interaction of SINE
462 RNA with rRNAs by PARIS2 and observed such RNA pairs. The sparsity of SINEUP–rRNA interactions may
463 be due to a limitation of psoralen cross-linking, which has bias for staggered uridines and may not access
464 RNA structures positioned inside complex ribonucleoprotein regions[55]. Alternatively, the efficiency of
465 PARIS2 to capture RNA–RNA interactions may be limited by the level of expression of RNAs as well as by
466 the brief half-lives of such interactions[41,55]. Nevertheless, our PARIS2 data identifies for the first time a
467 direct interaction of SINEUP-GFP (SINEB2) RNA and miniSINEUP-GFP (FRAM) RNA with both 18S and 28S

468 rRNA, which is consistent with previous polysome profiling data[38]. Even though we capture long RNA-RNA
469 interactions, in terms of sequence complementarity we unambiguously identify only short discontinuous
470 stretches of base-pairing between SINE-derived RNA and rRNA. However, this may still be sufficient to
471 influence translation, given that circular IRESs have 7-mer–long complementarity to 18S rRNA[43] and HCV
472 IRES needs only 3 nucleotide-long Watson-Crick base-pairing with 18S rRNA to induce translation[53].
473 Moreover, RNA–RNA interactions in 3D depend heavily on non-Watson-Crick base pairing and non-
474 covalent interactions, such as base-stacking and tertiary contacts by purine to minor-groove and tetraloop
475 to tetraloop receptor[64], which are not easily predicted. Therefore, we decided to take a closer look at the
476 SINE-rRNA interactions.

477 Notably, we discovered an interaction between a region comprising helix 14 and 15 of 18S rRNA (Int 1)
478 and motif 1 and 2 of SINEB2 RNA. FRAM RNA motifs 1 and 2 along with motifs 5 and 6 also interacted with
479 helix 15 of 18S rRNA. Helix 14 of 18S rRNA is known to interact with ribosomal protein rpL23[65]. In addition,
480 a protein ABCE1 (ATP Binding Cassette Subfamily E Member 1), which is involved in translation initiation
481 and termination, interacts with helix 14 during translation initiation[66]. Moreover, helices 14 and 15 bind
482 with eukaryotic initiation factor 5b (eIF5b), a ribosome-dependent GTPase that mediates joining of 40S
483 and 60S subunits of ribosomes[67]. This indicates the possibility that SINEUP motifs 1 and 2 are involved in
484 the joining of ribosomal subunits during translation initiation. Furthermore, we observed contact of the
485 region between motif 5 and 6 of SINEB2 RNA with domain III helices 55, 56, and 59 of 28S rRNA. Helix 59
486 is a part of the ribosomal tunnel exit site, from where the nascent peptide emerges out of the ribosome[68].
487 It suggests that SINEB2 RNA not only takes part in initiating target mRNA translation but also in elongation.
488 In addition, SINEB2 RNA motif 5 and the surrounding region was found to bind with domain I, helix 25,
489 expansion segments ES7c and ES7d of 28S rRNA. The connection with ES7c was also found to be conserved
490 in FRAM RNA, in which the SL1 region (containing a partial match to motif 5) participated in the
491 interaction. ES7 is the largest expansion segment positioned at the back of the 60S ribosomal subunit. This
492 highlights the importance of SINEUP structure motifs in their interaction with the large subunit of the
493 ribosome. We also observed interaction of FRAM RNA SL1 with 18S rRNA helices 21, 37, and 38. Because
494 the helix 38–42 region is crucial for mRNA binding and placement of anticodon arms for tRNAs[69], this
495 interaction is notable for the proper positioning of SINEUP target mRNA and translation initiation. FRAM
496 RNA also interacted with domains I, III, IV, and V of 28S rRNA.

497 These contacts at multiple rRNA sites indicate the dynamic changes in SINE RNA–rRNA interaction that
498 occur as translation progresses and highlight the role of SINEUPs in regulating both translation initiation

499 and elongation. Moreover, SINEUP RNAs were found to interact with other protein-coding RNAs and
500 lncRNAs (**Supplementary Table 4**). However, among these RNA partners only *Ank3* (Ankyrin 3) was
501 observed to interact with the SINEB2 RNA region. ANK3 protein links membrane proteins to spectrin-actin
502 cytoskeleton and is involved in neuronal development[70], but role of *Ank3* mRNA in translation is not
503 known. It is to be noted that although we captured the dynamics of SINE RNA structure and interaction
504 with rRNA and RBP, many of these processes are very transient in nature and it is unlikely to obtain a
505 complete view of all intermediate RNA structures and interaction kinetics in possibly very short time
506 windows by existing in vivo methods. Considering this, further 3D structure analysis of the tertiary
507 connections of SINEUP RNA with rRNA and translation complexes by using cryogenic electron microscopy
508 or other approaches is needed in the future to construct a full 3D model of larger molecular complexes.
509 In this study, the experimental 2D structure data-based prediction of 3D structure models for 18 SINE-
510 derived RNAs identified stabilizing stacking interactions within structure motifs and demonstrated
511 strikingly similar dynamics of mouse and human SINE-derived RNAs in 3D space in the subcellular
512 compartments. We list all the base-stacking and non-canonical interactions determined from these
513 models, which will provide a foundation for future experimental 3D structure analysis. In addition, the 3D
514 structure models of functionally inactive antisense *Gadd45α* SINEB2-a and related structure mutants
515 confirmed the similarity in global structure and SL1 of functional Δ27-36 mutant with antisense *Uchl1*
516 SINEB2 which was missing in the wild-type and other non-functional mutants. This further supports the
517 idea that such short structure motifs are crucial for non-coding RNA function.

518 In summary, this study offers a repertoire of the sequences, structures, and functional features of several
519 SINE-derived RNA elements for the first time and demonstrates that evolutionarily distant SINE RNAs have
520 a convergent translation upregulatory function driven by common short structure motifs. We show that
521 mouse and human functional SINE-derived RNAs follow a similar pattern of structural dynamics and SINE
522 RNA–RBP interactions are compartment-specific. Our study reveals direct interaction of mouse and
523 human SINE-derived transcripts with human small and large subunit of rRNAs, suggesting a mechanism
524 for SINEUP-mediated post-transcriptional gene regulation. This is the first study to provide a large number
525 of RNA 3D structure models of functional SINE-derived RNAs. Altogether the data presented here is a key
526 step towards filling the void in the structural and functional analysis of SINE-transcribed RNA and will
527 facilitate future studies that aim to resolve the SINEUP mechanism and biomolecular complexes involved.

528

529 **Data availability**

530 The icSHAPE sequencing data were deposited in Gene Expression Omnibus (GEO) with accession numbers
531 GSE146407 (for whole cell icSHAPE libraries of SINEUP-GFP) and GSE224534 (for whole cell icSHAPE
532 libraries of miniSINEUP-GFP and nuclear-cytoplasmic fractionated icSHAPE libraries), and GSE243220 (for
533 whole cell icSHAPE libraries of antisense *Gadd45α* SINEB2-a deletion mutants). The PARIS2 sequencing
534 data and SINEUP RNA modification data are accessible through GEO accession numbers GSE224533 and
535 GSE224018 respectively. Nuclear-cytoplasmic fractionated seCLIP sequencing data is available through
536 the GEO accession number GSE227250.

537

538 **Acknowledgements-**

539 We are grateful to the late Dr. Silvia Zucchelli (Department of Health Sciences, Center for Autoimmune
540 and Allergic Diseases (CAAD) and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD),
541 University of Piemonte Orientale, Novara, Italy) for her valuable suggestions and scientific discussion. This
542 work is dedicated to her memory. We are thankful to Dr. Howard Y. Chang from Standford University for
543 gifting us icSHAPE-probing reagent. We thank Dr. Byron Lee from Chang lab at Stanford University for
544 helpful suggestions for successful icSHAPE library preparation. We thank Souta Nagayama from P.
545 Carninci's lab for his help in deletion mutant analysis.

546

547 **Funding**

548 This work was supported by RIKEN Centre for Integrative Medical Sciences and RIKEN Centre for Life
549 Science Technology from the Japanese Ministry of Education, Culture, Sports, Science and Technology
550 (MEXT); and Basic Science and Platform Technology Program for Innovative Biological Medicine grant
551 from the Japan Agency for Medical Research and Development (AMED grant no.: 18am0301014 to PC).

552

553 *Conflict of interest statement.* S.G. and P.C. are inventors on patent US9353370B2 and related applications
554 in Europe, USA and Japan, and H.S., H.T., S.G., and P.C. are inventors on patent application
555 102018000002411 and the related application PCT/IB2019/050914 held by SISSA in Italy and TranSINE
556 Therapeutics in the UK, of which S.G. and P.C. are founders and H.T. owns shares.

557 H.S., N.T., M.V., H.T., and P.C. are inventors on the E.U. patent application EP3992289A1
558 (WO2022090733A1) held by TranSINE Therapeutics in the UK. S.G., N.T., H.T., and P.C. are inventors on
559 patent application WO2022064221A1 held by TranSINE Therapeutics in the UK. These conflicts of interest
560 were not associated with the funding of this work.

561

562 **Author Contributions**

563 H.S. H.T. and P.C wrote, reviewed, and edited the manuscript with comments from all the authors and
564 interpreted the data. H.S. designed and performed the SINEUP functional analysis, whole cell and nuclear-
565 cytoplasmic fractionated icSHAPE, icSHAPE for deletion mutants and analyzed respective data. H.S.
566 performed 2D and 3D structure modelling and analyzed PARIS data. M.N.Z.V. analyzed seCLIP data and
567 RNA modification data from Oxford Nanopore direct-RNA sequencing. N.T. performed seCLIP, RNA
568 modification experiments and prepared direct-RNA sequencing libraries. H.S.N. prepared PARIS libraries
569 and icSHAPE libraries for deletion mutants. H.T., P.C, and S.G. conceptualized and provided scientific
570 discussion. P.C. provided resources and secured funding for the project.

571

572 **Materials and Methods**

573 **Plasmids and cloning**

574 SINEUP-GFP (pcDNA3.1-based) synthetic SINEUP containing antisense *Uchl1* SINEB2[10] was used to
575 construct long SINEUP effector domain–swapping mutants. To swap the antisense *Uchl1* inverted SINEB2
576 of SINEUP-GFP with other SINEB2 sequences of interest, two restriction sites for SacII and ClaI restriction
577 enzymes were introduced at the start and end of the SINEB2. For this purpose, a QuikChange II Site-
578 Directed Mutagenesis Kit (Agilent, catalog #200523) was used. All the mutagenesis primers were designed
579 using QuikChange Primer Design Program (Agilent) and are listed in **Supplementary Table 5a**. The SINEB2
580 regions were PCR-amplified from RIKEN mouse cDNA clones AK143784 (antisense to *Txnip*) and AK054076
581 (antisense to *Gadd45α*), then digested by the SacII and ClaI restriction enzymes and sub-cloned in SINEUP-
582 GFP to create swapping mutants of antisense *Txnip* and antisense *Gadd45α* SINEB2-a and -b respectively.
583 Antisense *Gadd45α* SINEB2-ab and ba, and all the full constructs of antisense *Uxt* SINEB2–swapping
584 mutants were obtained from Genewiz. The highly active antisense *Uchl1*_Δ5'-32nt mutant[30] was included
585 in screening as a positive control. Direct and deleted antisense (ΔAS) *Uchl1* SINEB2 mutants were created
586 by a conventional restriction digestion, PCR, and cloning method, and the SINEUP mutant with deleted
587 binding domain (ΔBD) was purchased from TransSINE Technologies.
588
589 Previously described miniSINEUP-GFP[29] plasmid consisting of antisense *Uchl1* SINEB2 as the effector
590 domain and −40/+4 binding domain design with respect to the start codon of GFP mRNA was used as a
591 backbone for miniSINEUP effector-domain swapping. A conventional PCR amplification, restriction
592 digestion, ligation, and cloning strategy was used to generate miniSINEUP-GFP mutants shown in **Fig. 1e**.
593 The SINEB2 region of RIKEN mouse cDNA clones AK029359 (from 774–960 bp for antisense *Uxt-b* SINEB2),
594 AK032194 (from 769–982 bp for antisense *Nars2* SINEB2), AK041742 (from 2334–2533 bp for antisense
595 *Abhd11* SINEB2), and AK048309 (from 1203–1416 bp for antisense Epb4.9 SINEB2) was amplified by PCR
596 (all primer sequences are shown in **Supplementary Table 5b**). To prepare the effector domain for
597 antisense *Uchl1* SINEB2 + antisense *Uxt-b* SINEB2 (**Fig. 1e**), the SINEB2 regions of RIKEN cDNA clones
598 AK078321 (from 521–690 bp for antisense *Uchl1* SINEB2) and AK029359 (from 774–960 bp for antisense
599 *Uxt-b* SINEB2) were first amplified by PCR separately and then ligated to make one effector-domain insert
600 (all primer sequences are shown in **Supplementary Table 5b**). Similarly, in the case of 3× antisense *Uchl1*
601 SINEB2, PCR-amplified SINEB2 from AK078321 (from 521–690 bp) was consecutively ligated twice to
602 generate three repeats of the same SINEB2. The rest of the SINEB2 effector domains originating from
603 various mouse natural antisense lncRNAs (**Fig. 1e**, and **Supplementary Table 1**) and the consensus

604 sequence for SINEB2 subfamily B3 from the Repbase database[27,47] were commercially obtained from
605 Hokkaido System Science Co., Ltd. The backbone plasmid, miniSINEUP-GFP, and amplified/commercial
606 SINEB2 effector-domain inserts were digested by EcoRI and HindIII restriction enzymes, then ligated and
607 cloned.
608 The commercially available pEGFP-C2 (Clontech) plasmid was used as a sense EGFP construct. The
609 miniSINEUP-GFP FRAM plasmid design has been previously described[31].
610

611 **Cell culture and transfection**

612 Human embryonic kidney (HEK293T/17) cells (CRL-11268, ATCC) were grown in a 10-cm culture dish in
613 DMEM (1×) + GlutaMAX-1 (Thermo Fisher Scientific, 10569-044) supplemented with 10% fetal bovine
614 serum (Sigma, 172012) and 1% penicillin streptomycin solution (Wako, 168-23191) at 37°C and 5% CO_2
615 for 3–5 days. 70–90% confluent cells were treated with 0.05% w/v Trypsin-0.53 mmol/L EDTA 4Na Solution
616 with phenol red (Wako, 202-16931) for 5 min at 37°C to detach the cells, and 0.5×10^6 cells were seeded
617 in each well of 6-well plates (Falcon). After 24 h, cells were transfected with pEGFP-C2 and SINEUP-GFP
618 (long) or miniSINEUP-GFP swapping mutant plasmids at a molar ratio of 1:4.3 for long SINEUPs GFP [0.19
619 pmol pEGFP-C2 + 0.81 pmol of long SINEUP] and 1:4.6 for miniSINEUP-GFP [0.19 pmol pEGFP-C2 + 0.87
620 pmol of miniSINEUP-GFP] using 10 µL of Lipofectamine2000 (Thermo Fisher Scientific, 11668-019). After
621 24 h of transfection, cells were harvested.

622

623 **Protein extraction and Western blot**

624 Cells were washed with D-PBS (-) (Wako, 045-29795) and lysed using 140 µL of 10× Cell Lysis buffer (Cell
625 Signaling Technology, 9803S) with PMSF (Cell Signaling Technology, 8553S) per well in a 6-well plate.
626 Samples were rotated at low speed for 1 h at 4°C and then centrifuged (20,000 × *g*) for 10 min at 4°C. The
627 supernatant was collected, and the protein concentrations were measured by DC Protein Assay (BioRad,
628 500-0116). The absorbance (750 nm) was measured by Multimode Plate Reader ARVO X3 (PerkinElmer).
629 For Western blotting, 20 µg of extracted proteins mixed with 2× SDS sample buffer were separated on
630 10% SDS PAGE gel (Mini PROTEAN TGX Precast Gel, 10%, 12-well comb; BioRad, 456-1035) and transferred
631 to a nitrocellulose membrane (Amersham Protran Premium ECL 0.45 µm; GE Healthcare Life Sciences,
632 10600013) through semi-dry transfer at 25 V for 30 min using Trans-BLOT SD Semi-dry Transfer Cell
633 (BioRad). Tris-glycine buffer containing 20% methanol was used for transfer. The membranes were
634 blocked with 5% nonfat dry milk (Cell Signaling Technology, 9999S) in Tris Buffered Saline with Tween-20
635 (10× TBST, Cell Signaling Technology, 9997) for 30 min with shaking. Proteins were immunoblotted with
636 primary antibodies followed by horseradish peroxidase– (HRP) conjugated secondary antibodies. EGFP
637 was detected by anti-GFP rabbit serum A-6455 (Life Technologies), ACTINB was detected by monoclonal
638 anti-β-actin antibody (Sigma Aldrich, A5441). Protein signals were detected by ECL Western Blotting
639 Detection Reagent (GE Healthcare Life Sciences, RPN2109) with FUJI LAS-3000 system (Fujifilm) and
640 FUSION (Vilber-Lourmat). The band intensities were analyzed by Image J version 1.48 software (National
641 Institutes of Health).

642

643 **RNA extraction, cDNA synthesis, and quantitative RT-PCR**

644 HEK293T/17 cells were washed with D-PBS (-) (Wako, 045-29795) and detached using 0.05% w/v trypsin
645 (Wako, 202-16931). Cells were collected by centrifugation for 5 min (6000 × *g*) at 4°C, and total RNAs were
646 subsequently extracted using RNeasy Mini Kit (Qiagen, 74104). Extracted RNA samples were processed
647 through three rounds of DNase I digestion using a TURBO DNA-free Kit (Invitrogen, AM1907). RNA quality
648 was checked on Agilent 2100 Bioanalyzer (Agilent Technologies) using an Agilent RNA 6000 Nano Kit. Then,
649 500 ng of the total RNA was used for cDNA synthesis using the Prime Script 1st Strand cDNA Synthesis Kit
650 (TaKaRa, 6110A). Random 6-mer primers (0.3 µL) and 1.7 µL of oligodT primers were mixed to capture
651 RNAs. 2× diluted cDNA (1 µL) was used as a template for the quantitative real-time PCR (qRT-PCR). For
652 qRT-PCR, a SYBR Premix Ex Taq (Tli RNaseH Plus) Kit (TaKaRa, RR420S) was used. Each reaction contained
653 1 µL of cDNA, 0.4 µL of reverse primer (10 µM), 0.4 µL of forward primer (10 µM), 10 µL of SYBR premix
654 ExTaq, 0.4 µL of ROX reference dye (50×), and 10 µL of nuclease-free water (reaction volume was 20 µL).
655 EGFP and SINEUP-GFP levels were normalized by GAPDH. qRT-PCR was performed by StepOnePlus and
656 7900HT Real-Time PCR Systems (Applied Biosystems) under the following conditions: hold stage (1 cycle)
657 for 30 s at 95°C, cycling stage (40 cycles) for 5 s at 95°C and for 30 s at 60°C, melt curve stage. Reverse
658 transcription samples were taken to check for any plasmid DNA contamination in cDNA. Each sample was
659 performed in technical triplicates (n = 3, C_T standard deviation > 0.2) and with biological replicates (n ≥ 3).
660 2^-ΔΔCt ±SD was analyzed by StepOne v2.3 and SDS v2.4 software (Thermo Fisher Scientific). The qRT-
661 PCR primer sequences were as follows:

662 hGapdh_Fw: TCTCTGCTCCTCCTGTTC

663 hGapdh_Rv: GCCCAATACGACCAAATCC

664 EGFP_Fw: GCCCGACAACCACTACCTGAG

665 EGFP_Rv: CGGCGGTCACGAACTCCAG

666 SINEUP-GFP_Fw: CTGGTGTGTATTATCTCTTATG

667 SINEUP-GFP_Rv: CTCCCGAGTCTCTGTAGC

668 miniSINEUP-GFP_Fw: TTAATACGACTCACTATAGGGAGAC

669 miniSINEUP-GFP_Rev: CCGCATCTAGTGAACCGTCA

670

671 **Phylogenetic analysis and hierarchical clustering**

672 MEGA version 5 software[71] was used for sequence-based alignment, and mouse SINEB2 sequences tested
673 as effector domains in long and mini SINEUPs were aligned by ClustalW. SINEB2 sequence–based
674 molecular phylogenetic analysis was done using the Maximum Likelihood method. The evolutionary
675 history was inferred by the Maximum Likelihood method based on the Tamura 3-parameter model. Initial
676 tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ
677 algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood
678 approach and then selecting the topology with superior log likelihood value. A discrete Gamma
679 distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter =
680 4.1774)). The tree was drawn to scale, with branch lengths measured in the number of substitutions per
681 site.

682

683 **icSHAPE library preparation**

684 The icSHAPE standard protocol was followed[12,49]. Briefly, HEK293T/17 cells were co-transfected with
685 pEGFP-C2 and SINEUP-GFP or miniSINEUP-GFP swapping mutants using Lipofectamine2000 reagent in 6-
686 well plates. Cells were harvested after 24 h of transfection, pooled together, and treated with NAI-N_3
687 reagent (modified) (NAI-N_3 reagent was received as a gift from Dr. Howard Chang's lab at Stanford
688 University) for in vivo SHAPE modification or with DMSO (mock control). Two biological replicates for each
689 of the NAI-N_3 and DMSO libraries were made. RNA was extracted using RNeasy Mini Kit (Qiagen, 74104)
690 and digested with DNaseI to remove DNA contamination using a TURBO DNA-free Kit (Invitrogen,
691 AM1907). RNA was ribo-depleted using a magnetic RiboZero kit (Illumina, MRZH11124), and RNA quality
692 was checked by Agilent 2100 Bioanalyzer (Agilent Technologies) nano and pico chip. Next, 500 ng of ribo-
693 depleted RNA was biotin-clicked to label-probed bases followed by RNA fragmentation for 40 s using RNA
694 fragmentation reagents (Thermo Fisher Scientific, AM8740). Fragmented RNA was end-repaired, and the
695 icSHAPE adapter ligation standard protocol was followed[49] (adapter sequences in **Supplementary Table**
696 **6a**). Next, size selection was performed, and RNA was reverse-transcribed using icSHAPE-specific
697 barcoded reverse-transcription primer followed by streptavidin-based selection and size selection of
698 truncated cDNA products. Purified cDNA was circularized using CircLigase II enzyme (Epicentre, CL9025K),
699 then library PCR, size selection, and quantification were performed. Resultant libraries were sequenced
700 on a HiSeq2500 platform (Illumina, 100 bases, single-end). All the oligos used in the icSHAPE library
701 preparation are noted in **Supplementary Table 6a**.

702

703 **Comparison of icSHAPE profiles, secondary structure modelling, and motif analysis**

704 The standard icSHAPE bioinformatics pipeline was used[49]. Briefly, after removing PCR duplicates and
705 adapter trimming, raw sequencing data were mapped to EGFP and SINEUP transcripts index using
706 Bowtie2. Next, transcript abundance and reverse transcription (RT) stop numbers were calculated before
707 calculating the correlation between replicates and replicate merging. The mean of the RT stops of the 90–
708 95% most reactive bases was normalized as 1 and all other RT stops were scaled proportionally. Then
709 enrichment reactivity scores were calculated by subtracting the background signal (DMSO data), and
710 outliers were removed by 90% Winsorization (in which the top 5th percentile is set to 1 and the bottom
711 5th percentile is set to 0). Valid enrichment reactivity scores were filtered and visualized in integrated
712 genome viewer (IGV v2.11.1). The icSHAPE enrichment scores were used as a soft constraint in RNAfold
713 (ViennaRNA Package)[72]. Linear mapping method was used to derive pairing probabilities, and a method
714 described by Zarringhalam *et al.* (2012)[73] was used to incorporate guiding pseudo energies into the folding
715 algorithm. To draw the secondary structure, RNAcanvas[74] (https://rna2drawer.app/) was used.

716 Sequence and icSHAPE-guided 2D structure of all the SINEB2s were pairwise compared with the structure
717 and sequence of antisense *Uchl1* SINEB2 using ExpaRNA tool[50], which computes the exact matching
718 sequence-structure motifs common to two RNAs. Thus, derived motif files from all the pairwise
719 comparisons were then screened to extract the positions of Uchl1 SINEB2s that were present in one or
720 multiple motifs in all of the comparisons. Since SINEB2 sequences are highly variable, in most cases only
721 a partial sequence match could be found. Due to their repeating sequence composition, in some instances
722 one query sequence was matched to multiple motifs, or one motif matched to multiple positions in the

723 query. In that case, the longer matching stretch was preferred, and different motif numbers were assigned
724 based on their sequence and structure neighborhood and position on SINEB2. Additionally, SINEUP 2D
725 motifs were compared with existing 3D motifs in the RNA Characterization of Secondary Structure Motifs
726 database (CossMos)[51].

727

728 **Nuclear-cytoplasmic fractionation for icSHAPE**

729 HEK293T/17 cells were co-transfected with 0.8 pmol SINEUP-GFP or miniSINEUP-GFP FRAM (0.8 pmol)
730 and 0.2 pmol sense pEGFP-C2 plasmids (Clontech) in each well of a 6-well culture plate. Four 6-well plates
731 for each sample were prepared. At 24 h post-transfection, the medium was aspirated, and the cells were
732 washed and collected in 1 mL of 1× D-PBS (-) (Nacalai tesque, 14249-24) per well. Cells from one 6-well
733 plate (~1×10^7 cells) were pooled together in a 15-mL tube and centrifuged at 300 × *g* for 2 min at room
734 temperature. The resultant supernatant was removed, and the pelleted cells were resuspended in either
735 250 µL of 100 mM NAI-N3 reagent (Enamine, EN300-247313) for modification or in 5% DMSO (Sigma
736 Aldrich, D2650) solution in 1× PBS (Gibco) for mock treatment. The reaction was incubated on a rotating
737 wheel at 37°C for 5 min at room temperature and then stopped by centrifugation at 2500 × *g* for 1 min at
738 4°C and further removal of the supernatant. Next, the protocol by Pandya-Jones and Black (2009)[75] was
739 utilized for cellular fractionation. In brief, the NAI-N3- and DMSO-treated cell pellets were resuspended
740 in ice-cold NP-40 lysis buffer (10 mM Tris-HCl pH 7.5, 0.15% NP-40, 150 mM NaCl), then the lysate was
741 added on top of 2.5 volumes of chilled 25% sucrose cushion in lysis buffer without NP-40 followed by
742 centrifugation at 19,000 × *g* for 10 min at 4°C. The resultant supernatant containing the cytoplasmic
743 fraction was collected. The pellet was further gently rinsed with 200 µL chilled 1× PBS/1 mM EDTA and
744 centrifuged at 9,730 × *g* for 2 min at 4°C to remove the supernatant. Next, to extract the nuclear fraction,
745 the protocol by Brugiolo *et al.* (2017)[76] was followed. The pellet was resuspended in 100 µL nuclear buffer
746 I (50% glycerol v/v, 20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM NaCl, 0.85 mM DTT, 0.125 mM PMSF)
747 and then in 900 µL nuclear buffer IIA (20 mM HEPES pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 7.5
748 mM $MgCl_2$, 1M Urea, 1% NP-40, 400 U of RNase inhibitor). The solution was mixed thoroughly by vortexing
749 and then pipetting up and down, and then incubated on ice for 20 min. Finally, the nucleoplasmic fraction
750 was extracted by centrifuging at 14,000 × *g* for 5 min at 4°C and collecting the supernatant. The success
751 of fractionation was checked by Western blot and qRT-PCR for known nuclear and cytoplasmic marker
752 genes. Four nucleoplasmic or cytoplasmic fractions, each coming from four 6-well plates of the same
753 sample, were pooled together for further RNA extraction and icSHAPE library preparation as described
754 above. All the steps are the same as in the original protocol[49], except the removal of free 3' RNA linkers
755 after ligation, for which the completed ligation reaction was treated with 4 µL $RecJ_f$ (New England Biolabs,
756 M0264L) and 3 µL 5' Deadenylase (New England Biolabs, M0331S) at 37°C for 1 h and further cleaned up
757 by using a Zymo RNA Clean and Concentrator-5 purification kit (Zymo Research, R1016). All the oligos used
758 in the library preparation are noted in **Supplementary Table 6a**.

759

760 **seCLIP library preparation from nuclear-cytoplasmic fractionated samples**

761 Approximately 0.5×10^7 HEK293T/17 cells were plated on a 10-cm culture dish and after 24 h, transfected
762 with SINEUP-GFP and pEGFP-C2 plasmids as described above. After 24 h of transfection, 2×10^7 cells were
763 gently washed with 1× PBS and then crosslinked by 400 mJ/cm, 254 nm UV light irradiation on ice with

764 the lid off for 15 min. Next, the cells were collected by scraping in 1 mL chilled PBS followed by
765 centrifugation at 200 × *g* for 5 min at 4°C. The resultant supernatant was discarded, and the pellet was
766 used for nuclear-cytoplasmic fractionation following the protocol by Brugiolo *et al.* (2017)[76]. The pellet
767 was first resuspended in 2 mL hypotonic buffer (10 mM Tris-HCl 7.5 pH, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5
768 mM DTT, 1:200 proteinase inhibitor just prior to use) and incubated on ice for 5 min followed by
769 centrifugation at 500 × *g* for 10 min at 4°C. The supernatant containing hypotonic fraction was discarded,
770 and the pellet was resuspended in 1 mL lysis buffer 0.3 (50 mM Tris-HCl 7.5 pH, 150 mM NaCl, 2 mM
771 $MgCl_2$, 0.3% NP-40, 1:200 proteinase inhibitor just prior to use) and incubated on ice for 10 min. The mix
772 was then centrifuged at 1000 × *g* for 5 min at 4°C, and the supernatant containing the cytoplasmic fraction
773 was collected in a new 1.5-mL tube. The pellet was mixed with 1 mL lysis buffer 0.5 (50 mM Tris-HCl 7.5
774 pH, 150 mM NaCl, 2 mM $MgCl_2$, 0.5% NP-40, 1:200 proteinase inhibitor just prior to use) and incubated
775 on ice for 10 min followed by centrifugation at 1000 × *g* for 5 min at 4°C. The supernatant was discarded,
776 and the pellet was first resuspended in 100 µL nuclear buffer I (as used for icSHAPE fractionation, without
777 addition of PMSF) and then in 900 µL nuclear buffer IIA (as used for icSHAPE fractionation) with 1:200
778 proteinase inhibitor just prior to use. The pellet was mixed well by vortexing and then pipetting up and
779 down, and then incubated on ice for 15 min. The reaction mix was then centrifuged at 15,000 × *g* for 5
780 min at 4°C, and the nucleoplasm containing supernatant was transferred to a new 1.5-mL tube. All
781 fractions were separately centrifuged at 20,000 × *g* for 5 min at 4°C and tested for the presence of markers
782 specific to nuclear or cytoplasmic fractions by Western blot.

783 Next, the seCLIP protocol described by Van Nostrand *et al*. (2017)[46] was utilized to prepare nuclear and
784 cytoplasmic seCLIP libraries. In brief, nucleoplasmic and cytoplasmic lysates were treated with 4 U of
785 DNase I and 10 µL of 1:25 diluted RNase I (100 U/µL) in a thermomixer at 37°C and 1200 rpm for 5 min.
786 The reaction was then cooled on ice, 7 µL Murine RNase Inhibitor (New England Biolabs, M0314) was
787 added, and the mix was centrifuged at 15,000 × *g* for 15 min at 4°C. The resultant supernatant was mixed
788 with 15 µg of anti-hnRNPK antibody (Abcam, ab39975) coupled to 200 µL magnetic beads (Dynabeads M-
789 280 Sheep Anti-mouse IgG, Invitrogen, 11201D) and incubated at 4°C overnight. Around 2% of the input
790 samples (containing all RNA-protein complexes) were saved before immunoprecipitation.
791 Immunoprecipitated samples were first end-treated with FastAP thermo-sensitive alkaline phosphatase
792 (Thermo Scientific, EF0651) and T4 poly nucleotide kinase (New England Biolabs, M0201L) and then
793 ligated with 3' RNA linker (with sample barcodes, see sequence in **Supplementary Table 6b**). Next, both
794 immunoprecipitated and input RBP–RNA complexes were visualized by Western blot imaging and the
795 region corresponding to HNRNPK and above (from 55–180 kDa) was sliced from the preparative
796 nitrocellulose membrane blot. The RNA was extracted by treating the membrane slices with urea and
797 proteinase K followed by acid phenol–chloroform extraction and purification by using a Zymo RNA Kit
798 (R1013). Further treatment of samples and cDNA library preparation are the same as described in the
799 original protocol[46]. Libraries were sequenced on the HiSeq2500 platform (150 cycles, single-end).

800

801 **Analysis of nuclear-cytoplasmic fractionated seCLIP data**

802 Fastq files were processed with two rounds of cutadapt (which trims off adapters and controls for double
803 ligation events) before being mapped to a custom genome (hg38 + SINEUP construct) using STAR v2.7.8a
804 with parameters as in the ENCODE eCLIP project pipeline (https://www.encodeproject.org/eclip/)
805 (namely, --outFilterMultimapNmax 1, --outFilterMultimapScoreRange 1, --outFilterScoreMin 10, --

806 alignEndsType EndToEnd). Reads mapped on the sense strand of the SINEUP construct were removed.
807 Peaks were called on mapped reads using clipper v0.2.0 (https://github.com/YeoLab/clipper) with custom
808 data files (gff, exons bed file) containing the SINEUP region, with default parameters. Scripts from the
809 merge_peaks github repository (https://github.com/YeoLab/merge_peaks) were used to normalize
810 seCLIP reads over input reads for each replicate (overlap_peakfi_with_bam.pl) before merging input-
811 normalized peaks into a final set of significant binding regions
812 (compress_l2foldenrpeakfi_for_replicate_overlapping_bedformat.pl).

813

814 **Preparation of chemically modified SINEUP-GFP RNA transcribed in vitro**

815 Transcribed SINEUP-GFP RNA was synthesized in vitro as described by Toki *et al.* (2020)[77] by using a
816 mMESSAGE mMACHINE SP6 Transcription Kit (Thermo Fisher Scientific) and was then modified by
817 following the protocol by Mandal and Rossi (2013)[78]. For the 20% Ψ modification, UTP was replaced with
818 20% pseudouridine-5'-triphosphate (TriLink, USA; final concentration, 7.5 mM) by mixing the reagents
819 with 40 ng/µL of linearized SINEUP RNA. To add a poly-A tail, 1–10 µg of RNA transcribed in vitro was
820 treated with Escherichia coli poly(A) polymerase (5000 U/mL, New England Biolabs, M0276) at 37°C for
821 30 min followed by clean-up using an RNeasy Mini kit (Qiagen) and the modified transcripts of SINEUP-
822 GFP RNA were recovered.

823

824 **Extraction of modified SINEUP-GFP RNA transcribed in-cell**

825 HEK293T/17 cells were plated on a 10-cm dish and SINEUP-GFP plasmids were transfected for 24 h; then
826 2×10^7 cells were harvested by Trizol/chloroform extraction, and total RNAs containing SINEUP-GFP RNAs
827 at the aqueous phase were extracted by using an RNeasy mini Kit (Qiagen). The RNA was then digested
828 with 4 U of TURBO DNase (TURBO DNA-free kit, Invitrogen, AM1907) at 37°C for 90 min followed by DNase
829 inactivation.

830

831 **Pull-down of SINEUP-GFP RNAs transcribed in-cell**

832 The total RNA volume was adjusted to 100 µL with lysis buffer (50 mM Tris-HCl pH 7.0, 10 mM EDTA, 1%
833 SDS, 1:100 RNase inhibitor) and mixed with two volumes of hybridization buffer (750 mM NaCl, 50 mM
834 Tris-HCl pH 7.0, 1 mM EDTA, 1% SDS, 15% formamide, 1:200 RNase inhibitor) with additional 6.6%
835 formamide. Then, 100 pmol of SINEUP-GFP-specific total probe (2 µL of 50 µM probe) was added to this
836 mix (probe design as described by Toki *et al.*, 2020)[38], and the reaction was incubated at 37°C overnight
837 with vertical rotation. Next, 100 µL of MagCapture Tamavidin 2-REV magnetic beads (WAKO) were washed
838 and resuspended in 100 µL lysis buffer. This bead suspension was added to the completed hybridization
839 reaction and incubated at 37°C for 30 min with mixing. The solution was then magnetically separated, and
840 the supernatant was discarded. The beads were washed four times with pre-warmed wash buffer (2×
841 saline sodium citrate, 0.5% SDS, 1:200 RNase inhibitor), and the supernatant was discarded. After this, the
842 beads were resuspended in 100 µL ProK buffer (10 mM Tris-HCl pH 7.5, 0.8% SDS, 15 mM EDTA) without
843 proteinase K and incubated at 65°C for 5 min with end-to-end mixing at 450 rpm. The first pull-down RNA

844 was then extracted by the Trizol/chloroform method and purified by miRNeasy mini column (Qiagen,
845 217004). This pull-down and RNA extraction was repeated once again to enrich SINEUP-GFP RNAs.

846

847 **Library preparation of Oxford Nanopore direct-RNA sequencing**

848 Libraries of the direct SINEUP-GFP RNA transcribed in vitro and in-cell were prepared with an SQK-RNA002
849 kit (Oxford Nanopore Technology). The original reverse-transcription adapter (RTA) was used for in vitro
850 samples and four barcoded RTA adapters described by Leger *et al.* (2021)[79] were used for in-cell samples.
851 The libraries were applied to MinION Mk1C sequencer (Oxford Nanopore Technologies) and sequenced
852 for 72 h.

853

854 **Modification detection using long-read sequencing**

855 Raw fast5 files from direct-RNA sequencing of modified SINEUP RNA transcribed in-cell and in vitro were
856 base-called using guppy v5.0.7 with RNA-centric options (--trim_strategy rna, --u_substitution true, --
857 reverse_sequence true). After preparing reference transcriptome annotation files (bed and fasta
858 annotation files from bedparse gtf2bed and bedtools getfasta, respectively), the base-called data were
859 mapped to the reference using minimap2 v2.17. To detect modified k-mers in the data we followed the
860 nanocompore pipeline (https://github.com/tleonardi/nanocompore/). Briefly, this involved indexing the
861 raw data, followed by running nanopolish eventalign to realign the raw signal-level data to the k-mers of
862 the reference. After this, the signal was collapsed to the k-mer level using NanopolishComp, and
863 nanocompore was used to compare the sample transcribed in-cell with the different transcription
864 conditions in vitro. Results of this comparison were filtered by both log odds ratio (absolute value > 1) and
865 *p* value (<0.01). Finally, a peak-calling algorithm
866 (https://gist.github.com/tleonardi/0bb31e6a380e5766f04f4e197d36b38e) was used to get rid of multiple
867 adjacent significant sites, because a single modified nucleotide can often cause multiple neighbouring k-
868 mers to be measured as significant sites, when in reality there is just one modification. The peak-calling
869 algorithm merges these adjacent sites into a single k-mer, which helps to highlight true-positive sites.

870

871 **PARIS library preparation**

872 HEK293T/17 cells were transfected with pEGFP-C2 and SINEUP-GFP (SINEB2) or miniSINEUP-GFP FRAM
873 plasmids in 6-well plates as described above. After 24 h of transfection, cells were gently washed with 1×
874 PBS and treated with 150 µL 1× PBS (control) or 1 mg/mL Amotosalen (MedChemExpress, HY-107004A)
875 crosslinking solution added to each well of the 6-well plates, followed by incubation under standard cell-
876 culture conditions (37°C, 5% CO_2) for 30 min. Next, cells were irradiated for 30 min by 365-nm UV light
877 while on ice with swirling of the plate every 10 min. Then, the crosslinking solution was removed, and the
878 cells were scraped off into 1 mL chilled 1× PBS in each well and transferred to a 15-mL tube. Crosslinked
879 cells were pelleted by centrifuging at 300 × *g* for 5 min at 4°C, and the supernatant was discarded. Next,
880 for the total nucleic acid (TNA) extraction and PARIS library preparation, the protocol by Zhang *et al.*
881 (2021)[41] was used with some modifications. Briefly, the cell pellet was lysed by 6 M GuSCN with pipetting
882 followed by treatment with 0.5 M EDTA and passed through a 25G needle 20 times. After this, the proteins

883 were digested by proteinase K (final concentration 1 mg/mL) at 37°C for 1 h on a thermomixer with
884 shaking, and solution was again passed through the 25G needle. After TNA was precipitated by phenol-
885 isopropanol precipitation as described in Zhang *et al.* (2021)[41], 100 µg of TNA sample was treated with 50
886 units of TURBO DNase (Invitrogen, AM2239) at 37°C for 20 min. RNA was precipitated using phenol-
887 isopropanol and further purified using RNA clean and concentrator kit (Zymo Research, R1015). The RNA
888 quality was measured with Bioanalyzer Nano Kit (Agilent). DNase-treated RNA (10 µg) was fragmented by
889 using 5 units of ShortCut RNase III (New England Biolabs, M0245) for 5 min at 37°C and the reaction was
890 stopped by adding 10 µL nuclease free water, 100 µL of RNA binding buffer (Zymo Research, R1015) and
891 150 µL of 100% ethanol and then purified through a column from the purification kit (Zymo Research,
892 R1015) followed by quality check with a Bioanalyzer small RNA kit (Agilent).

893 For 2D gel purification of RNA duplexes, 8% 1-mm thick denatured first-dimension gel was prepared using
894 the UreaGel System (National Diagnostics, EC-833). A 10-µL sample of fragmented RNA was run on the
895 first-dimension gel in 0.5x Novex TBE running buffer (Invitrogen, LC6675) at 30 W for 9 min. The gel was
896 visualized under 300 nm transillumination after staining with SYBR Gold (Invitrogen, S11494) and each
897 lane was sliced between 50 bp to topside from the first-dimension gel. Next, 16% 1-mm thick denatured
898 second-dimension gel was prepared using the UreaGel System (National Diagnostics, EC-833). The second-
899 dimension gel was poured to encapsulate the gel slices from the first-dimension gel, and 0.5x TBE buffer
900 warmed to 60°C was used to facilitate the denaturing of crosslinked RNA. The gel was run at 30 W for 50
901 min and then stained with SYBR Gold, and crosslinked RNA was sliced out from the upper diagonal of the
902 2D gel. The gel slices were cut into smaller pieces and transferred to a 0.5-mL tube with a hole drilled in
903 the bottom with a 20G needle. This 0.5-mL tube was placed inside a 2-mL tube and centrifuged at 16,000
904 x *g* for 5 min to crush the gel through the hole. Then 1.5 mL of crushing buffer (20 mM Tris-HCl pH 7.5,
905 0.25 M sodium acetate, 1 mM EDTA, 0.25% SDS) was added to the crushed gel and rotated at room
906 temperature overnight. After this, 500 µL of gel slurry was transferred to a Micro Spin column (Cytiva,
907 27356501) and centrifuged at 3400 x *g* for 1 min or until all gel slurry was filtered. This filtered gel was
908 then passed through an Amicon Ultra 10K column (Millipore, UFC501096) and further purified using the
909 Zymo RNA Clean and Concentrator-5 column.

910 Next, in order to perform proximity ligation, 10 µL of the purified double-stranded RNA was treated with
911 7.5 unit/µL of T4 RNA Ligase I, high concentration (New England Biolabs, M0437) in a reaction mix
912 containing 1× T4 RNA Ligase Buffer (New England Biolabs, M0437M), 1U SUPERase•In RNase inhibitor
913 (Invitrogen, AM2694), and 0.5 mM ATP at room temperature overnight. The proximity-ligated RNA was
914 purified by using a Zymo RNA clean and concentrator-5 column.

915 To reverse crosslink the RNA, the samples were mixed with 3 µL of 25 mM acridine orange on ice and
916 irradiated with 254-nm UV light for 30 min. The reverse crosslinked sample was transferred to a new 1.5-
917 mL tube and purified using a Zymo RNA Clean and Concentrator-5 column. Further, the reverse crosslinked
918 RNA was heated at 80°C for 90 s and then snap-cooled on ice. The adapter was ligated by adding 14 µL of
919 adapter ligation mix (1.5 units/µL T4 RNA ligase I high concentration, 1.5 µM adapter (5′ Biotin), 10%
920 DMSO, 12.5% PEG 8000, 5 mM DTT, 1× T4 RNA Ligase buffer) to 6 µL of reverse crosslinked RNA and
921 incubating the reaction for 3 h at room temperature. The sequences of all the oligos used in the PARIS
922 library preparation are listed in **Supplementary Table 6c**. After adapter ligation, free adapters were
923 removed by adding 3 units of $RecJ_f$ (New England Biolabs, M0264) and 1.7 units of 5′ deadenylase (NEB,
924 M0331S) and incubating at 37°C for 1 h. The adapter-ligated RNA was purified using a Zymo RNA Clean
925 and Concentrator-5 column and eluted in 10 µL of nuclease free water. Next, 2 µL of custom reverse

926 transcription primer (with barcode, 1 µM, sequence in **Supplementary Table 6c**) and 1 µL of 10 mM dNTPs
927 were added to the purified RNA and the sample was heated at 65°C for 5 min in a thermal cycler followed
928 by rapid cooling on ice. To this mix, 8 µL reverse transcriptase mix was added, which contained 4 µL of 5x
929 SSIV buffer (Mn2+) (250 mM Tris-HCl, 375 mM CH3COOK, 7.5mM $MnCl_2$), 10 mM DTT, 1 unit/µL
930 SUPERaseIn RNase inhibitor (Invitrogen, AM2694), 10 units/µL SuperScript IV (Invitrogen, 18090010). The
931 reaction was incubated at 25°C for 15 min, 42°C for 10 h; hold at 10°C.

932 The reverse transcribed RNA was captured on streptavidin-coated magnetic beads. For this purpose, 10
933 µL of Dynabeads M-270 Streptavidin (Invitrogen, 65305) were washed in 20 µL LiCl buffer (7 M LiCl, 20
934 mM Tris-HCl (pH7.5), 2 mM EDTA, 0.1% Tween20) and then resuspended in 45 µL of LiCl buffer. This beads
935 suspension was added to the reverse transcribed RNA and incubated at 37°C for 30 min with pipette
936 mixing in between. The supernatant was removed by separation on a magnetic stand, and beads were
937 washed three times with 100 µL TE wash buffer (10 mM Tris-HCl (pH7.5), 1 mM EDTA, 0.1% Tween20) and
938 finally resuspended in 30 µL elution buffer (1x RNaseONE buffer, 1% Tween20, 10 µM P3-tall primer).
939 Next, the beads solution was heated at 95°C for 5 min then immediately cooled on ice and magnetically
940 separated to recover the released cDNA containing supernatant in a new 1.5-mL tube. The beads were
941 again resuspended in 30 µL elution buffer, separated on a magnet, and supernatant was recovered and
942 transferred to the released cDNA tube from the previous step. Next, the cDNA was treated with 1 µL of
943 RNaseONE (Promega, M4261), 1 µL of RNaseH (TaKaRa, 2150A), and 1 µL of RNase cocktail (Invitrogen,
944 AM2286) for 30 min at 37°C and further purified by using a Zymo DNA Clean and Concentrator column
945 (Zymo Research, D4033).

946 After this, 4 µL of circularization reaction mix containing 5 units/µL CircLigase II (Lucigen, CL9021K), 1×
947 CircLigase II buffer (Lucigen, CL9021K), and 2.5 mM $MnCl_2$ was added to 16 µL of cDNA and incubated at
948 60°C for 100 min followed by 80°C for 10 min. Circularized cDNA (1 µL) was used to determine the PCR
949 cycles required for the library amplification. For this, 39 µL of qPCR mix containing 1× Phusion mix (New
950 England Biolabs, M0531S), 0.125 µM P6 Tall v4 primer, 0.125 µM P3 Tall v4 primer, 0.15 µL ROX, 1/3800
951 SYBR (Invitrogen, S-7563) was added to 1 µL circularized cDNA and amplified at 95°C for 45 s; 20 cycles of
952 95°C for 15 s, 65°C for 30 s, 72°C for 45 s on a StepOnePlus Real-Time PCR System (Applied Biosystems,
953 4376598). The qPCR was stopped just before the amplification plot reached the plateau and the
954 corresponding cycle number was noted. For the final library amplification, 6 µL of circularized cDNA was
955 added to 44 µL PCR mix (1× Phusion mix, 0.15 µM P6 Tall v4 primer, 0.15 µM P3 Tall v4 primer) and
956 amplified at 95°C for 45 s; pre-determined cycles of 95°C for 15 s, 65°C for 30 s, 72°C for 45 s on a routine
957 thermocycler. The PCR product was purified using Zymo DNA clean and concentrator column (Zymo
958 Research, D4013) and ran on 6% native TBE gel. DNA from 175 to 300 bp was excised, crushed, and
959 purified. Library quality was checked by using a Bioanalyzer High-sensitivity DNA chip and concentration
960 was quantified by using a qPCR KAPA library quantification kit (KAPABiosystems, KK4835). Four barcoded
961 libraries were pooled together and sequenced on HiSeq2500 platform (150 nt, single-end, rapid mode).

962

963 **PARIS data analysis**

964 To check the SINEUP regions interacting with rRNAs, the PARIS2 analysis strategy was used as described
965 by Zhang *et al.* (2021)[41] (https://github.com/minjiezhang-usc/PARIS2). In short, the 3' end adapters were
966 removed by Trimmomatic v0.39, and the sequencing reads were split according to the sample barcodes
967 using splitFastq script from icSHAPE pipeline[49]. Next, PCR duplicates were removed using the readCollapse

968 script from the icSHAPE pipeline[49] followed by removal of the 5' header using Trimmomatic v0.39. The
969 reads were then mapped to a custom genome generated by adding SINEUP-GFP (SINEB2), miniSINEUP-
970 GFP (FRAM), and sense EGFP genes as separate chromosomes to the hg38 masked genome[41], which were
971 previously engineered to contain a single copy of multicopy genes such as rRNA and snRNAs. Mapping
972 was done by STAR v2.7.10a following the parameters described in the PARIS2 paper[41]. To extract SINEUP-
973 rRNA interactions, first the reads that were mapped to rRNA were extracted by samtools (as in the PARIS2
974 analysis strategy) and then rRNA interacting reads in which one of the arms of the RNA-rRNA pair was
975 mapped to SINEUPs were extracted by using the grep command. The extracted file was sorted by samtools
976 and visualized on the integrated genome viewer (IGV v2.11.1). VfoldMCPX[80] was used to predict the
977 structure of the RNA-rRNA pair. Human 18S rRNA structure (ID: URS00005A14E2) and 28S rRNA structure
978 (ID: URS0000C873C2) were taken from the RNAcentral database (https://rnacentral.org/).

979

980 **3D structure modelling**

981 SINE RNAs 2D structure information derived from icSHAPE was machine translated into 3D structure
982 models by RNAComposer[81] (https://rnacomposer.cs.put.poznan.pl/). The 3D models were visualized by
983 using Mol* tool[82] of RCSB PDB[83] (https://www.rcsb.org/) and analyzed by RNApdbee 2.0[84]
984 (http://rnapdbee.cs.put.poznan.pl/) to recognize the positions with base-stacking and non-canonical
985 interactions.

986

987 **References**

988

989 1. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet* **30**, 121-3
990 (2014).
991 2. Novikova, I.V., Hennelly, S.P. & Sanbonmatsu, K.Y. Structural architecture of the human long
992 non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* **40**, 5034-51 (2012).
993 3. Somarowthu, S. et al. HOTAIR forms an intricate and modular secondary structure. *Mol Cell* **58**,
994 353-61 (2015).
995 4. Xue, Z. et al. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription
996 Factor to Specify the Cardiovascular Lineage. *Mol Cell* **64**, 37-50 (2016).
997 5. Guttman, M. & Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**,
998 339-46 (2012).
999 6. Johnson, R. & Guigo, R. The RIDL hypothesis: transposable elements as functional domains of
1000 long noncoding RNAs. *RNA* **20**, 959-76 (2014).
1001 7. Wilusz, J.E. et al. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A)
1002 tails. *Genes Dev* **26**, 2392-407 (2012).
1003 8. Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T. & Steitz, J.A. Formation of triple-helical
1004 structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad*
1005 *Sci U S A* **109**, 19202-7 (2012).
1006 9. Uroda, T. et al. Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53
1007 Pathway. *Mol Cell* **75**, 982-995 e9 (2019).
1008 10. Carrieri, C. et al. Long non-coding antisense RNA controls Uchl1 translation through an
1009 embedded SINEB2 repeat. *Nature* **491**, 454-7 (2012).

1010 11. Podbevsek, P. et al. Structural determinants of the SINE B2 element embedded in the long non-
1011 coding RNA activator of translation AS Uchl1. *Sci Rep* **8**, 3189 (2018).
1012 12. Ohyama, T. et al. An NMR-based approach reveals the core structure of the functional domain of
1013 SINEUP lncRNAs. *Nucleic Acids Res* **48**, 9346-9360 (2020).
1014 13. Kramerov, D.A. & Vassetzky, N.S. Short retroposons in eukaryotic genomes. *Int Rev Cytol* **247**,
1015 165-221 (2005).
1016 14. Ohshima, K. & Okada, N. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail.
1017 *Cytogenet Genome Res* **110**, 475-90 (2005).
1018 15. Su, M., Han, D., Boyd-Kirkup, J., Yu, X. & Han, J.J. Evolution of Alu elements toward enhancers.
1019 *Cell Rep* **7**, 376-385 (2014).
1020 16. Bourque, G. et al. Evolution of the mammalian transcription factor binding repertoire via
1021 transposable elements. *Genome Res* **18**, 1752-62 (2008).
1022 17. Sundaram, V. et al. Widespread contribution of transposable elements to the innovation of gene
1023 regulatory networks. *Genome Res* **24**, 1963-76 (2014).
1024 18. Roy-Engel, A.M. et al. Human retroelements may introduce intragenic polyadenylation signals.
1025 *Cytogenet Genome Res* **110**, 365-71 (2005).
1026 19. Policarpi, C. et al. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. *Cell Rep* **21**, 2879-
1027 2894 (2017).
1028 20. Zhang, X.O., Gingeras, T.R. & Weng, Z. Genome-wide analysis of polymerase III-transcribed Alu
1029 elements suggests cell-type-specific enhancer function. *Genome Res* **29**, 1402-1414 (2019).
1030 21. Faulkner, G.J. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet*
1031 **41**, 563-71 (2009).
1032 22. Fort, A. et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role
1033 for retrotransposons in pluripotency maintenance. *Nat Genet* **46**, 558-66 (2014).
1034 23. Lubelsky, Y. & Ulitsky, I. Sequences enriched in Alu repeats drive nuclear localization of long
1035 RNAs in human cells. *Nature* **555**, 107-111 (2018).
1036 24. Gong, C. & Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with
1037 3' UTRs via Alu elements. *Nature* **470**, 284-8 (2011).
1038 25. Mouse Genome Sequencing, C. et al. Initial sequencing and comparative analysis of the mouse
1039 genome. *Nature* **420**, 520-62 (2002).
1040 26. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome*
1041 *Res* **110**, 462-7 (2005).
1042 27. Bao, W., Kojima, K.K. & Kohany, O. Repbase Update, a database of repetitive elements in
1043 eukaryotic genomes. *Mob DNA* **6**, 11 (2015).
1044 28. Karijolich, J., Zhao, Y., Alla, R. & Glaunsinger, B. Genome-wide mapping of infection-induced
1045 SINE RNAs reveals a role in selective mRNA export. *Nucleic Acids Res* **45**, 6194-6208 (2017).
1046 29. Zucchelli, S. et al. SINEUPs are modular antisense long non-coding RNAs that increase synthesis
1047 of target proteins in cells. *Front Cell Neurosci* **9**, 174 (2015).
1048 30. Takahashi, H. et al. Identification of functional features of synthetic SINEUPs, antisense lncRNAs
1049 that specifically enhance protein translation. *PLoS One* **13**, e0183229 (2018).
1050 31. Schein, A., Zucchelli, S., Kauppinen, S., Gustincich, S. & Carninci, P. Identification of antisense
1051 long noncoding RNAs that function as SINEUPs in human cells. *Sci Rep* **6**, 33605 (2016).
1052 32. Sun, L. et al. RNA structure maps across mammalian cellular compartments. *Nat Struct Mol Biol*
1053 **26**, 322-330 (2019).
1054 33. Kierzek, E. et al. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic*
1055 *Acids Res* **42**, 3492-501 (2014).
1056 34. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated
1057 pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).

1058 35. Lin, T.Y., Mehta, R. & Glatt, S. Pseudouridines in RNAs: switching atoms means shifting
1059 paradigms. *FEBS Lett* **595**, 2310-2322 (2021).
1060 36. Rinn, J.L. et al. Functional demarcation of active and silent chromatin domains in human HOX
1061 loci by noncoding RNAs. *Cell* **129**, 1311-23 (2007).
1062 37. Tsai, M.C. et al. Long noncoding RNA as modular scaffold of histone modification complexes.
1063 *Science* **329**, 689-93 (2010).
1064 38. Toki, N. et al. SINEUP long non-coding RNA acts via PTBP1 and HNRNPK to promote translational
1065 initiation assemblies. *Nucleic Acids Res* **48**, 11626-11644 (2020).
1066 39. (!!! INVALID CITATION !!! 38).
1067 40. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B.J. Global Mapping of Human RNA-RNA
1068 Interactions. *Mol Cell* **62**, 618-26 (2016).
1069 41. Zhang, M. et al. Optimized photochemistry enables efficient analysis of dynamic RNA
1070 structuromes and interactomes in genetic and infectious diseases. *Nat Commun* **12**, 2344
1071 (2021).
1072 42. Meng, Z., Jackson, N.L., Shcherbakov, O.D., Choi, H. & Blume, S.W. The human IGF1R IRES likely
1073 operates through a Shine-Dalgarno-like interaction with the G961 loop (E-site) of the 18S rRNA
1074 and is kinetically modulated by a naturally polymorphic polyU loop. *J Cell Biochem* **110**, 531-44
1075 (2010).
1076 43. Chen, C.K. et al. Structured elements drive extensive circular RNA translation. *Mol Cell* **81**, 4300-
1077 4318 e13 (2021).
1078 44. Reisser, S., Zucchelli, S., Gustincich, S. & Bussi, G. Conformational ensembles of an RNA hairpin
1079 using molecular dynamics and sparse NMR data. *Nucleic Acids Res* **48**, 1164-1174 (2020).
1080 45. Spitale, R.C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**,
1081 486-90 (2015).
1082 46. Van Nostrand, E.L. et al. Robust, Cost-Effective Profiling of RNA Binding Protein Targets with
1083 Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). *Methods Mol Biol* **1648**,
1084 177-200 (2017).
1085 47. Kalb, V.F., Glasser, S., King, D. & Lingrel, J.B. A cluster of repetitive elements within a 700 base
1086 pair region in the mouse genome. *Nucleic Acids Res* **11**, 2177-84 (1983).
1087 48. Lunyak, V.V. & Atallah, M. Genomic relationship between SINE retrotransposons, Pol III-Pol II
1088 transcription, and chromatin organization: the journey from junk to jewel. *Biochem Cell Biol* **89**,
1089 495-504 (2011).
1090 49. Flynn, R.A. et al. Transcriptome-wide interrogation of RNA secondary structure in living cells
1091 with icSHAPE. *Nat Protoc* **11**, 273-90 (2016).
1092 50. Heyne, S., Will, S., Beckstette, M. & Backofen, R. Lightweight comparison of RNAs based on
1093 exact sequence-structure matches. *Bioinformatics* **25**, 2095-102 (2009).
1094 51. Vanegas, P.L. et al. RNA CoSSMos: Characterization of Secondary Structure Motifs--a searchable
1095 database of secondary structure motifs in RNA three-dimensional structures. *Nucleic Acids Res*
1096 **40**, D439-44 (2012).
1097 52. Spahn, C.M. et al. Cryo-EM visualization of a viral internal ribosome entry site bound to human
1098 ribosomes: the IRES functions as an RNA-based translation factor. *Cell* **118**, 465-75 (2004).
1099 53. Matsuda, D. & Mauro, V.P. Base pairing between hepatitis C virus RNA and 18S rRNA is required
1100 for IRES-dependent translation initiation in vivo. *Proc Natl Acad Sci U S A* **111**, 15385-9 (2014).
1101 54. Lu, Z., Gong, J. & Zhang, Q.C. PARIS: Psoralen Analysis of RNA Interactions and Structures with
1102 High Throughput and Resolution. *Methods Mol Biol* **1649**, 59-84 (2018).
1103 55. Lu, Z. & Chang, H.Y. The RNA Base-Pairing Problem and Base-Pairing Solutions. *Cold Spring Harb*
1104 *Perspect Biol* **10**(2018).

1105 56. Vicens, Q. & Kieft, J.S. Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad*
1106 *Sci U S A* **119**, e2112677119 (2022).
1107 57. Vassetzky, N.S. & Kramerov, D.A. SINEBase: a database and tool for SINE analysis. *Nucleic Acids*
1108 *Res* **41**, D83-9 (2013).
1109 58. Indrieri, A. et al. Synthetic long non-coding RNAs [SINEUPs] rescue defective gene expression in
1110 vivo. *Sci Rep* **6**, 27315 (2016).
1111 59. Bon, C. et al. SINEUP non-coding RNAs rescue defective frataxin expression and activity in a
1112 cellular model of Friedreich's Ataxia. *Nucleic Acids Res* **47**, 10728-10743 (2019).
1113 60. Espinoza, S. et al. SINEUP Non-coding RNA Targeting GDNF Rescues Motor Deficits and
1114 Neurodegeneration in a Mouse Model of Parkinson's Disease. *Mol Ther* **28**, 642-652 (2020).
1115 61. Kramerov, D.A. & Vassetzky, N.S. Origin and evolution of SINEs in eukaryotic genomes. *Heredity*
1116 *(Edinb)* **107**, 487-95 (2011).
1117 62. Davis, D.R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23**, 5020-6 (1995).
1118 63. Evans, J.R. et al. Members of the poly (rC) binding protein family stimulate the activity of the c-
1119 myc internal ribosome entry segment in vitro and in vivo. *Oncogene* **22**, 8012-20 (2003).
1120 64. Bou-Nader, C. & Zhang, J. Structural Insights into RNA Dimerization: Motifs, Interfaces and
1121 Functions. *Molecules* **25**(2020).
1122 65. Spahn, C.M. et al. Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-
1123 ribosome and subunit-subunit interactions. *Cell* **107**, 373-86 (2001).
1124 66. Simonetti, A., Guca, E., Bochler, A., Kuhn, L. & Hashem, Y. Structural Insights into the
1125 Mammalian Late-Stage Initiation Complexes. *Cell Rep* **31**, 107497 (2020).
1126 67. Unbehaun, A. et al. Position of eukaryotic initiation factor eIF5B on the 80S ribosome mapped
1127 by directed hydroxyl radical probing. *EMBO J* **26**, 3109-23 (2007).
1128 68. Armache, J.P. et al. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome
1129 at 5.5-A resolution. *Proc Natl Acad Sci U S A* **107**, 19748-53 (2010).
1130 69. Heuer, A. et al. Cryo-EM structure of a late pre-40S ribosomal subunit from Saccharomyces
1131 cerevisiae. *Elife* **6**(2017).
1132 70. Kloth, K. et al. ANK3 related neurodevelopmental disorders: expanding the spectrum of
1133 heterozygous loss-of-function variants. *Neurogenetics* **22**, 263-269 (2021).
1134 71. Tamura, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
1135 evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9 (2011).
1136 72. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
1137 73. Zarringhalam, K., Meyer, M.M., Dotu, I., Chuang, J.H. & Clote, P. Integrating chemical
1138 footprinting data into RNA secondary structure prediction. *PLoS One* **7**, e45160 (2012).
1139 74. Johnson, P.Z., Kasprzak, W.K., Shapiro, B.A. & Simon, A.E. RNA2Drawer: geometrically strict
1140 drawing of nucleic acid structures with graphical structure editing and highlighting of
1141 complementary subsequences. *RNA Biol* **16**, 1667-1671 (2019).
1142 75. Pandya-Jones, A. & Black, D.L. Co-transcriptional splicing of constitutive and alternative exons.
1143 *RNA* **15**, 1896-908 (2009).
1144 76. Brugiolo, M., Botti, V., Liu, N., Muller-McNicoll, M. & Neugebauer, K.M. Fractionation iCLIP
1145 detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm
1146 and cytoplasm. *Nucleic Acids Res* **45**, 10452-10465 (2017).
1147 77. Toki, N., Takahashi, H., Zucchelli, S., Gustincich, S. & Carninci, P. Synthetic in vitro transcribed
1148 lncRNAs (SINEUPs) with chemical modifications enhance target mRNA translation. *FEBS Lett* **594**,
1149 4357-4369 (2020).
1150 78. Mandal, P.K. & Rossi, D.J. Reprogramming human fibroblasts to pluripotency using modified
1151 mRNA. *Nat Protoc* **8**, 568-82 (2013).

1152 79. Leger, A. et al. RNA modifications detection by comparative Nanopore direct RNA sequencing.
1153 *Nat Commun* **12**, 7198 (2021).
1154 80. Zhang, S., Cheng, Y., Guo, P. & Chen, S.J. VfoldMCPX: predicting multistrand RNA complexes.
1155 *RNA* **28**, 596-608 (2022).
1156 81. Biesiada, M., Purzycka, K.J., Szachniuk, M., Blazewicz, J. & Adamiak, R.W. Automated RNA 3D
1157 Structure Prediction with RNAComposer. *Methods Mol Biol* **1490**, 199-215 (2016).
1158 82. Sehnal, D. et al. Mol* Viewer: modern web app for 3D visualization and analysis of large
1159 biomolecular structures. *Nucleic Acids Res* **49**, W431-W437 (2021).
1160 83. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).
1161 84. Zok, T. et al. RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res*
1162 **46**, W30-W35 (2018).

1163

1164

1165

1166

1167

1168

1169

1170

1171

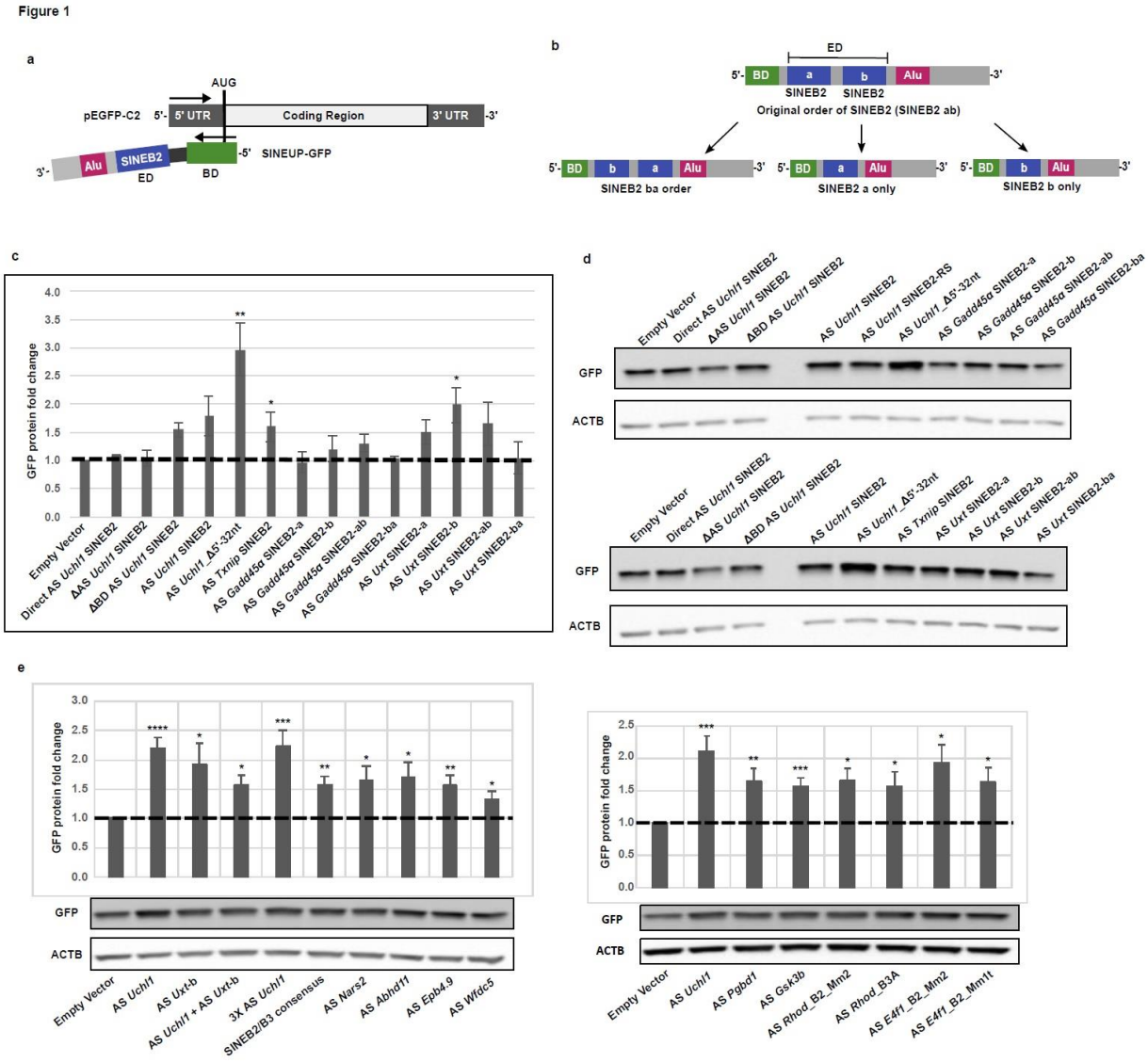


Figure 1. SINEB2 RNAs from different mouse antisense lncRNAs can act as SINEUP effector domain and upregulate GFP protein expression in HEK293T/17. (**a**) Schematic representation of the sense/antisense EGFP system used to validate the effect of different SINEB2 RNAs on translation. The SINEUP binding domain (BD) is shown in green. The inverted SINEB2 region is the effector domain (ED) shown in blue. Alu repeat and 3' tail of SINEUP are shown in magenta and gray respectively. (**b**) SINEUP construct design to verify independent and combined effects of two SINEB2 elements from the same antisense lncRNA. (**c**) Western blot–based EGFP band intensities normalized to ACTB expression as a measurement of SINEUP-induced protein fold change observed in the HEK293T/17 cell line at 24 h post-transfection. Empty vector, Direct AS Uchl1 SINEB2, and ΔAS Uchl1 SINEB2 (SINEB2 deletion) wereused as controls. (**d**) Western blot images with anti-GFP and anti-ACTB antibodies. ACTB was used as a loading control. Abbreviations: AS, antisense; Direct, direct orientation of SINEB2; Δ, deletion; a and b, two different SINEB2 elements from the same antisense lncRNA; ab and ba, the two SINEB2 elements in original and reversed order, respectively; AS Uchl1 SINEB2–RS, antisense *Uchl1* SINEB2 construct without any restriction enzyme sites around the SINEB2 elements. (**e**) GFP protein fold change in HEK293T cells after co-transfection with sense

1187 EGFP and miniSINEUP (at 24 h post-transfection). Western blot images and corresponding GFP band
1188 intensities (normalized to ACTB expression level) calculated using ImageJ software. An empty vector was
1189 used as the negative control and antisense *Uchl1* SINEB2 containing miniSINEUP (shown as AS *Uchl1*) as
1190 the positive control. Sample labels indicate names of antisense lncRNAs from which the SINEB2 sequences
1191 were isolated, which represent four sub-families (based on RepeatMasker annotation): B3, B3A, Mm2,
1192 and Mm1t. SINEB2 elements are from the B3 sub-family unless specifically stated. AS *Uchl1* + AS *Uxt*-b is
1193 the combination of AS *Uchl1* SINEB2 and the second SINEB2 of AS *Uxt*; 3× indicates three repeats of
1194 SINEB2; and SINEB2/B3 consensus is the B3 sub-family consensus sequence taken from the RepBase
1195 database. N = 3 (for c) and n = 5 (for e); error bars ± standard error mean; **** P < 0.00005; *** P <
1196 0.0005; ** P < 0.005; * P < 0.05 by two-tailed Student's *t*-test.

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207

1208

1209

1210

1211

1212

1213

1214

1215

1216

1217

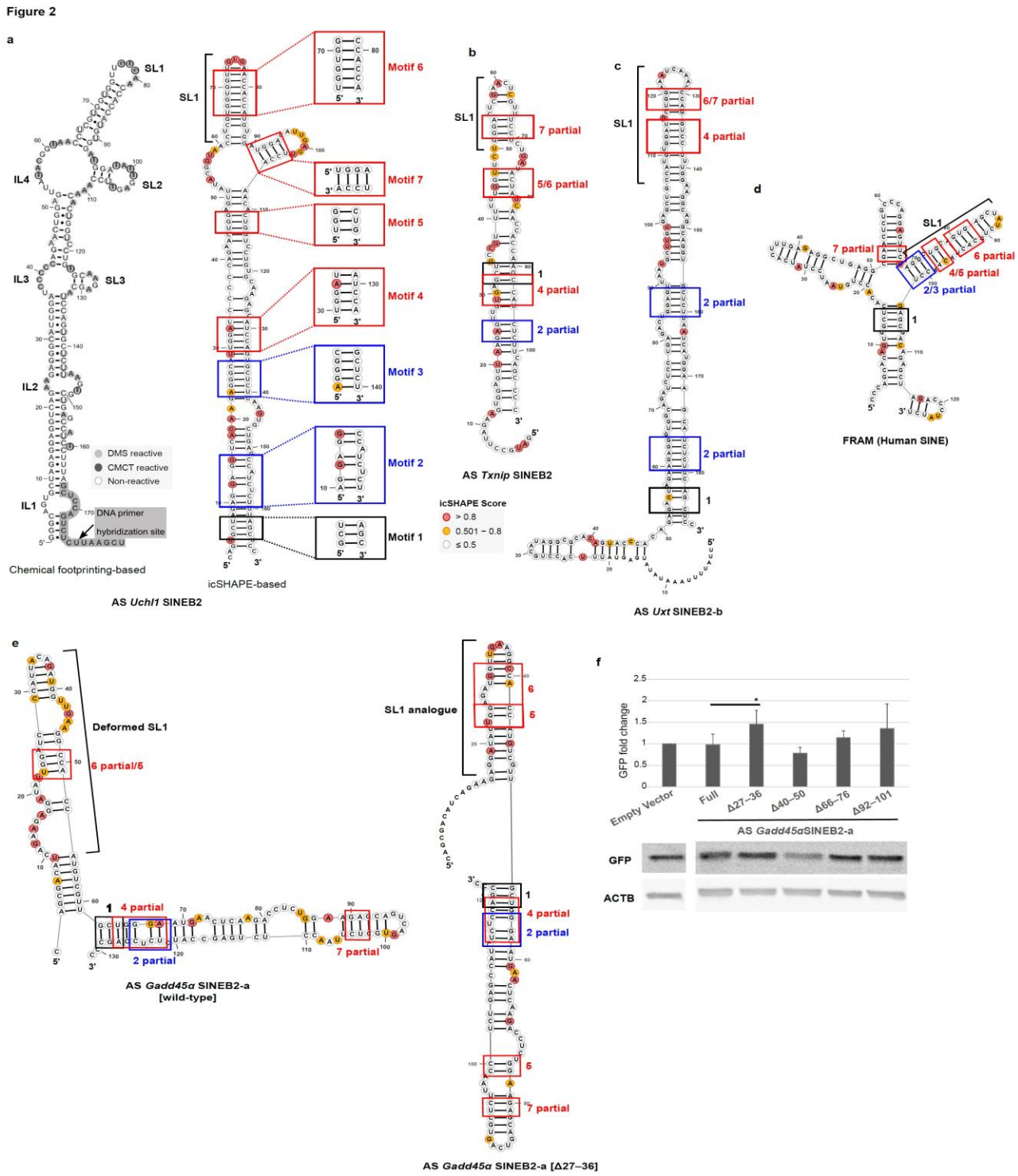


Figure 2. In-cell SINE structures and conserved motifs in functionally active SINEUPs. (**a**) Antisense (AS) *Uchl1* SINEB2 RNA secondary structure based on icSHAPE data (right) is compared with the one previously identified using chemical footprinting[11]. In chemical footprinting-based structure, DNA primer hybridization site is highlighted in gray and indicated by an arrow; light and dark gray circles represent DMS and CMCT reactive bases, respectively; non-reactive bases are outlined by a circle. In icSHAPE-based structure, squares highlight the common sequences and structure motifs in in-cell structures of SINE, with

1225 their corresponding motif number. AGG (blue) and UGG (red) motifs are shown. (**b**–**d**) Location of SINEUP
1226 motifs in icSHAPE-derived structures of functionally active SINE RNAs from (**b**) antisense *Txnip* SINEB2, (**c**)
1227 antisense *Uxt* SINEB2-b and, (**d**) human SINE element FRAM. (**e**) icSHAPE-derived RNA secondary
1228 structures and local SINEUP motifs (colors as in Figure 2a) of functionally dormant wild-type antisense
1229 *Gadd45α* SINEB2-a and its deletion mutant antisense *Gadd45α* SINEB2-a [Δ27–36] with gained SL1. (**f**)
1230 Effect of antisense *Gadd45α* SINEB2-a deletion mutants on GFP protein expression as measured by
1231 Western blot. An empty vector was used as the negative control. Four mutants with random deletion, as
1232 listed in the Figure, were compared with the full-length construct (Full). N = 3, error bars ± standard
1233 deviation, * $P < 0.05$ by two-tailed Student's *t*-test. AS, antisense; Δ, deletion; SL1, stem-loop 1. Nucleotide
1234 color indicates normalized icSHAPE reactivity score: red, ≥ 0.8; yellow, from above 0.5 to 0.8; white
1235 encircled with grey, ≤ 0.5; bases not encircled, no icSHAPE data available.

1236

1237

1238

1239

1240

1241

1242

1243

1244

1245

1246

1247

1248

1249

1250

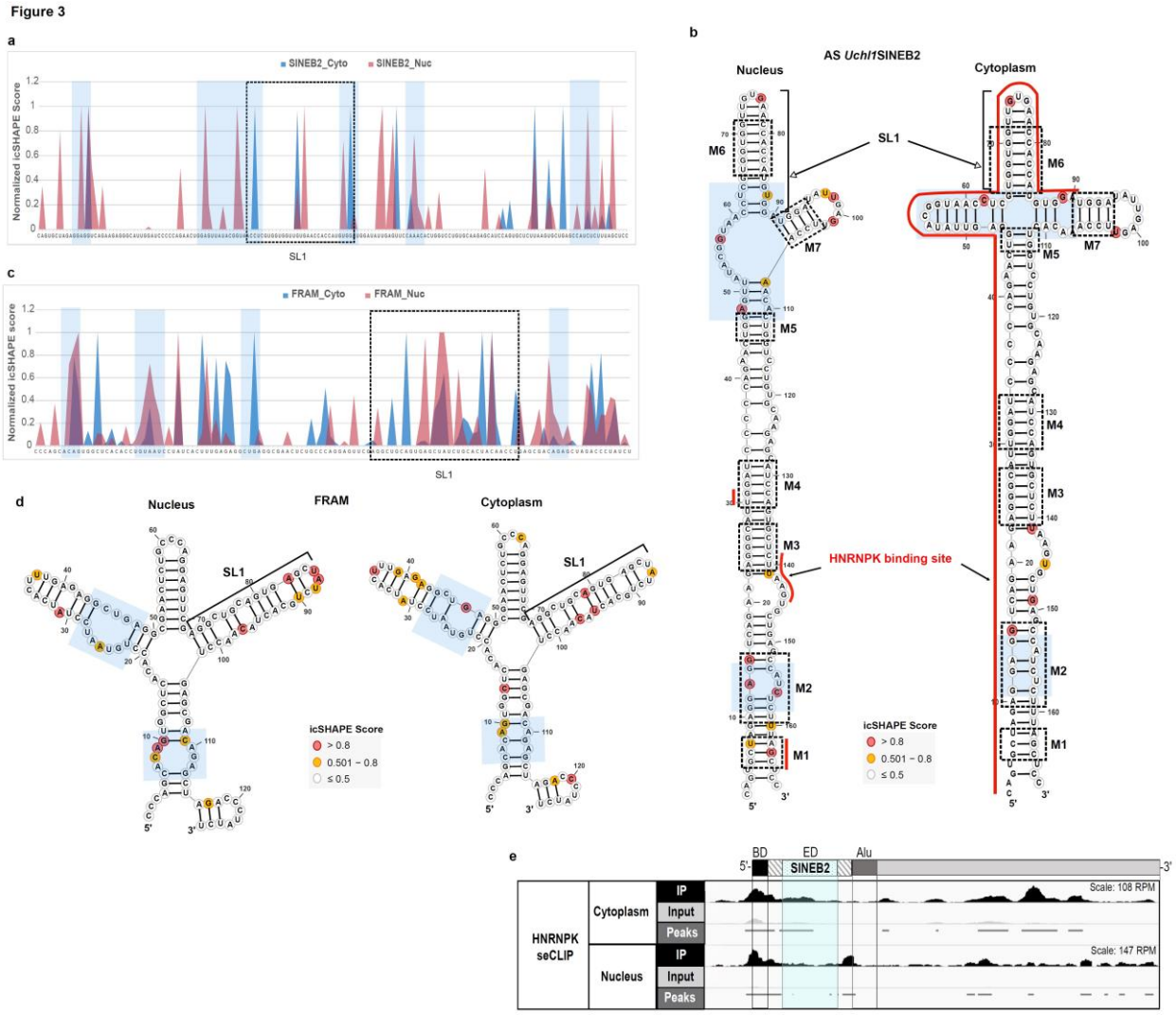


Figure 3. SINEUP structure dynamics and interaction with HNRNPK in the nucleus and cytoplasm. (**a**) Normalized icSHAPE reactivity score profiles of mouse antisense (AS) *Uchl1* SINEB2 RNA in the nucleus and cytoplasm and (**b**) related 2D structure models created from icSHAPE data. SINEUP motifs from M1 to M7 are indicated by dashed squares. (**c**) Normalized icSHAPE reactivity score profiles of human SINE RNA FRAM in the nucleus and cytoplasm and (**d**) their respective 2D structure models. Different structural regions between the compartments are shaded in blue boxes. (**e**) seCLIP data for HNRNPK binding to SINEUP-GFP RNA (of AS *Uchl1* SINEB2) in the nuclear and cytoplasmic fractions of HEK293T cells. IP: immunoprecipitation against HNRNPK; input: not immunoprecipitated, contains all RNA-protein complexes in the given fraction. Normalized peak regions representing significant enrichment of HNRNPK binding site in IP compared to input are indicated as horizontal bars. The SINEB2 region is highlighted in blue. BD, binding domain; ED, effector domain. Identified HNRNPK binding sites are marked by a red line on 2D structure models of AS *Uchl1* SINEB2 in panel b.

Figure 4

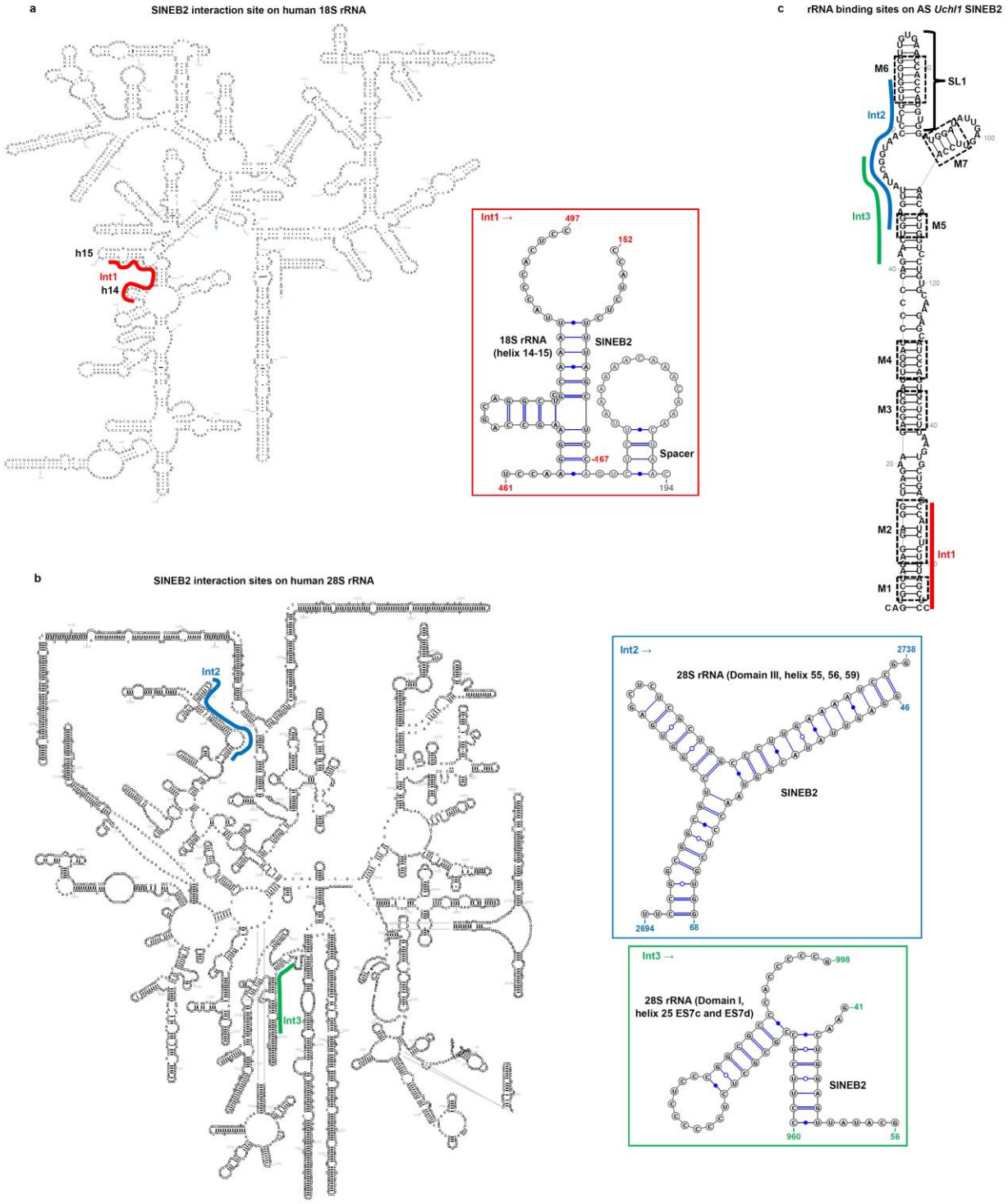


Figure 4. Mouse SINEB2 RNA interacts with human rRNA in cell. Antisense (AS) *Uchl1* SINEB2 interacting regions (Int) in human (**a**) 18S rRNA (Int1) and (**b**) 28S rRNA (Int2 and 3) identified by PARIS2. Different interaction sites are color-coded, and rRNA-SINEB2 pairs are shown in the corresponding color. Possible

1268 base-pairing based on sequence complementarity of interacting regions is shown. (**c**) SINEB2 regions
1269 interacting with 18S and 28S rRNA regions marked in (a) and (b) on icSHAPE data-driven SINEB2 structure
1270 from whole-cell lysate. SINEUP motifs from M1 to M7 are indicated by dashed squares.

1271

1272

1273

1274

1275

1276

1277

1278

1279

1280

1281

1282

1283

1284

1285

1286

1287

1288

1289

1290

1291

1292

1293

1294

1295

1296

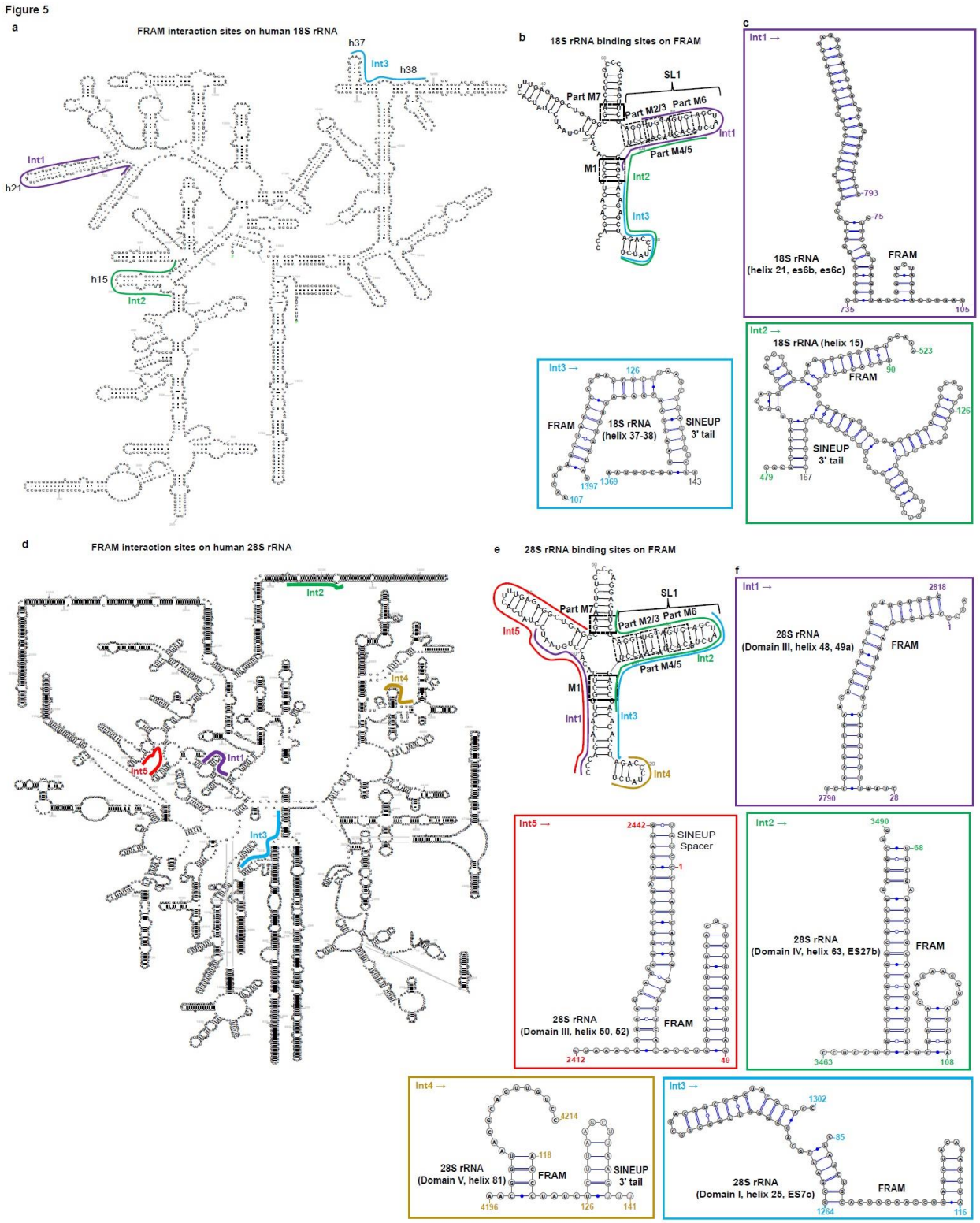


Figure 5. In-cell interaction of functional human SINE element FRAM with rRNA. (**a**) Regions of 18S rRNA found to pair with the SINEUP FRAM region in PARIS2 data. (**b**) Corresponding binding regions to (a) are highlighted on FRAM RNA. (**c**) Description of FRAM-18S rRNA pairs and interaction sites marked in (a) and

1301 (b). (**d**) 28S rRNA regions identified as FRAM interaction sites by PARIS2 and (**e**) their corresponding
1302 binding regions on FRAM. (**f**) Details of FRAM-28S rRNA interactions highlighted in (d) and (e). Different
1303 interaction sites (Int) are represented by different colors. Possible base-pairing based on sequence
1304 complementarity of RNA-RNA pairs is shown in (c) and (f). In (b) and (e) the icSHAPE-guided FRAM
1305 structure from whole cell lysate is shown. Matching regions to SINEUP motifs from M1 to M7 are indicated
1306 by dashed squares.

1307

1308

1309

1310

1311

1312

1313

1314

1315

1316

1317

1318

1319

1320

1321

1322

1323

1324

1325

1326

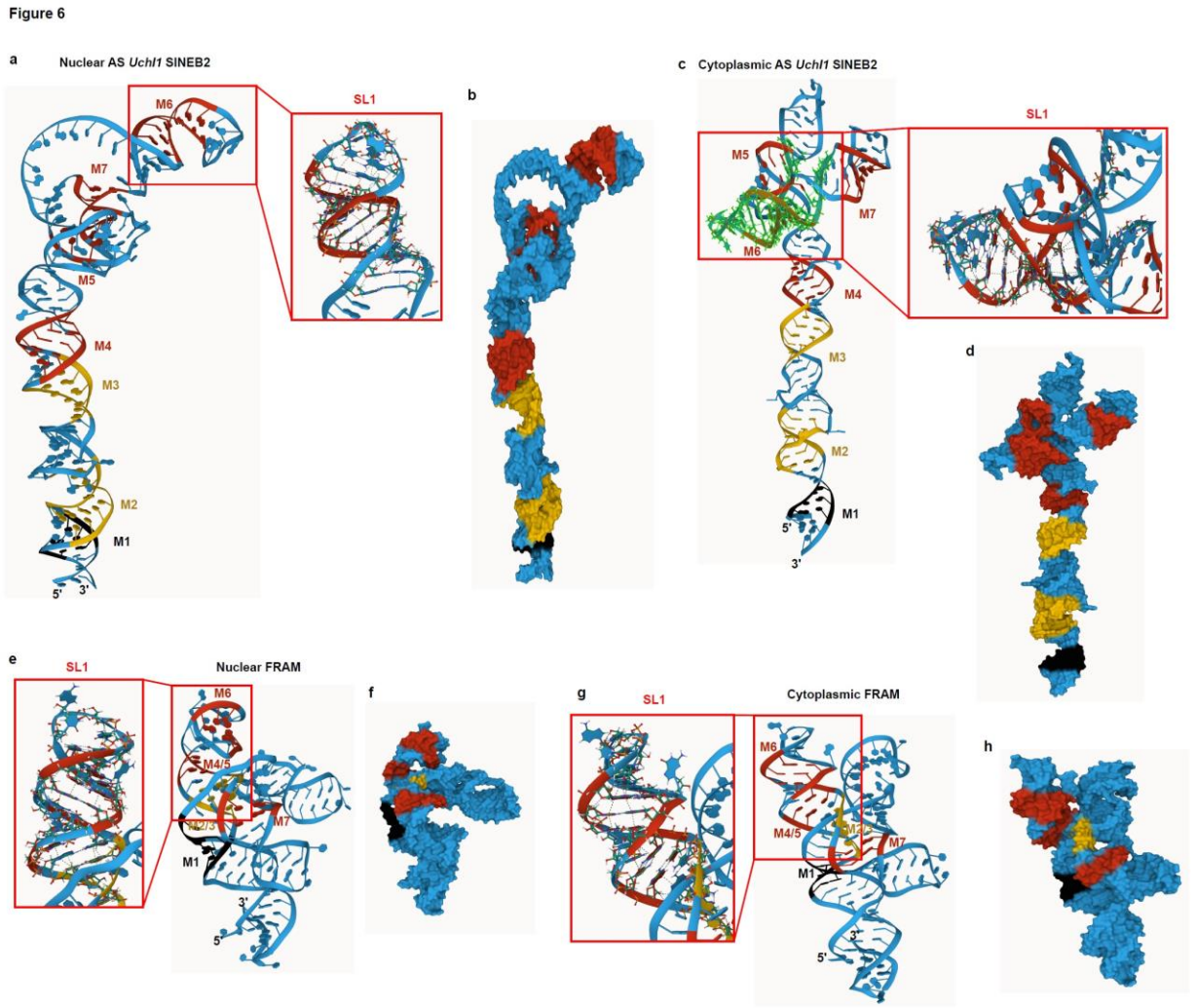
1327

1328

1329

1330

1331



1332

1333 **Figure 6. Machine-translated 3D RNA structure models based on in vivo SINE RNA 2D structures.** 3D
1334 structure models for (**a**–**d**) mouse antisense (AS) *Uchl1* SINEB2 RNA in the nucleus (**a**,**b**) and cytoplasm
1335 (**c**,**d**) and for (**e**–**h**) FRAM RNA in the nucleus (**e**,**f**) and cytoplasm (**g**,**h**). SINEUP structure motifs M1–M7
1336 are shown. Red frames show enlarged views of the SL1 region. Panels b, d, f, and h show the molecular
1337 surface view of the 3D models.

1338

1339

1340

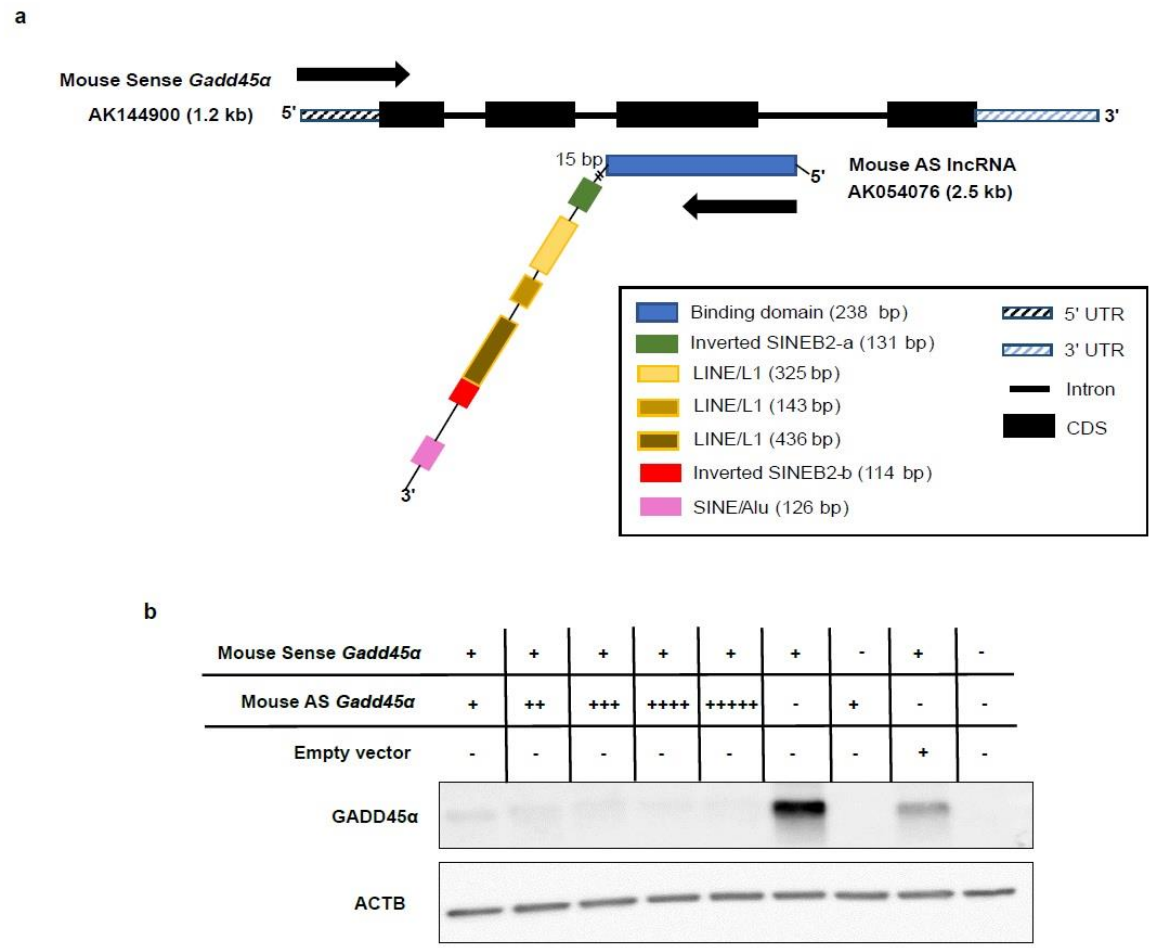
1341

1342

1343

1344

Extended Data Figure 1



1345

1346 **Extended Data Figure 1. Effect of mouse natural antisense *Gadd45α* lncRNA on sense *Gadd45α* protein**
1347 **expression.** (**a**) Molecular features of mouse sense/antisense *Gadd45α* RNAs. Antisense (AS) *Gadd45α*
1348 contains a sequence complementary to intron 2, exon 3, and intron 3 of sense Gadd45α mRNA,two
1349 inverted SINEB2s, three LINE, and one SINE-Alu elements. (**b**) Western blot analysis of sense Gadd45α
1350 protein expression in HEK293T/17 cells co-transfected with sense/antisense *Gadd45α* for 24 h. Empty
1351 vector, negative control; plasmid transfection at + 0.7 pmol; ++ 0.8 pmol; +++ 0.9 pmol; ++++ 1.0 pmol; -
1352 not transfected. ACTB is the loading control.

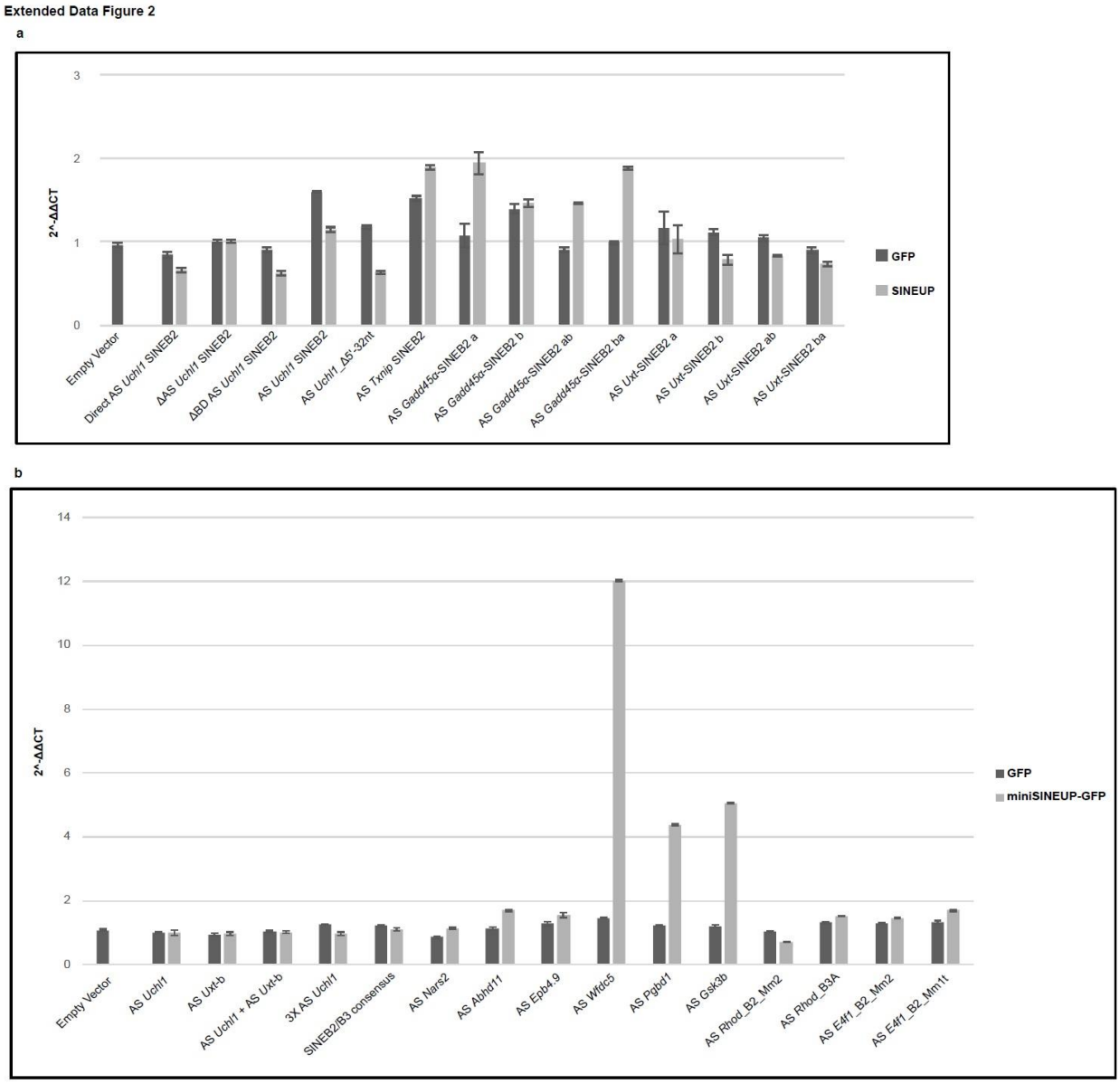
1353

1354

1355

1356

1357



1358

1359 **Extended Data Figure 2. qRT-PCR results for GFP mRNA (black bars) and SINEUP-GFP RNA (grey bars)**
1360 **expression.** Expression values are normalized to human GAPDH mRNA. Data were analyzed by using the
1361 $\Delta\Delta C_T$ method. RNA was extracted from (**a**) long SINEUPs (see description of terms in Figure 1c), with ΔAS
1362 *Uchl1* SINEB2 = 1 and from (**b**) MiniSINEUP-GFP (described in Figure 2b), with AS *Uchl1* = 1. Error bars are
1363 ± standard error mean for three biological replicates. AS, antisense.

1364

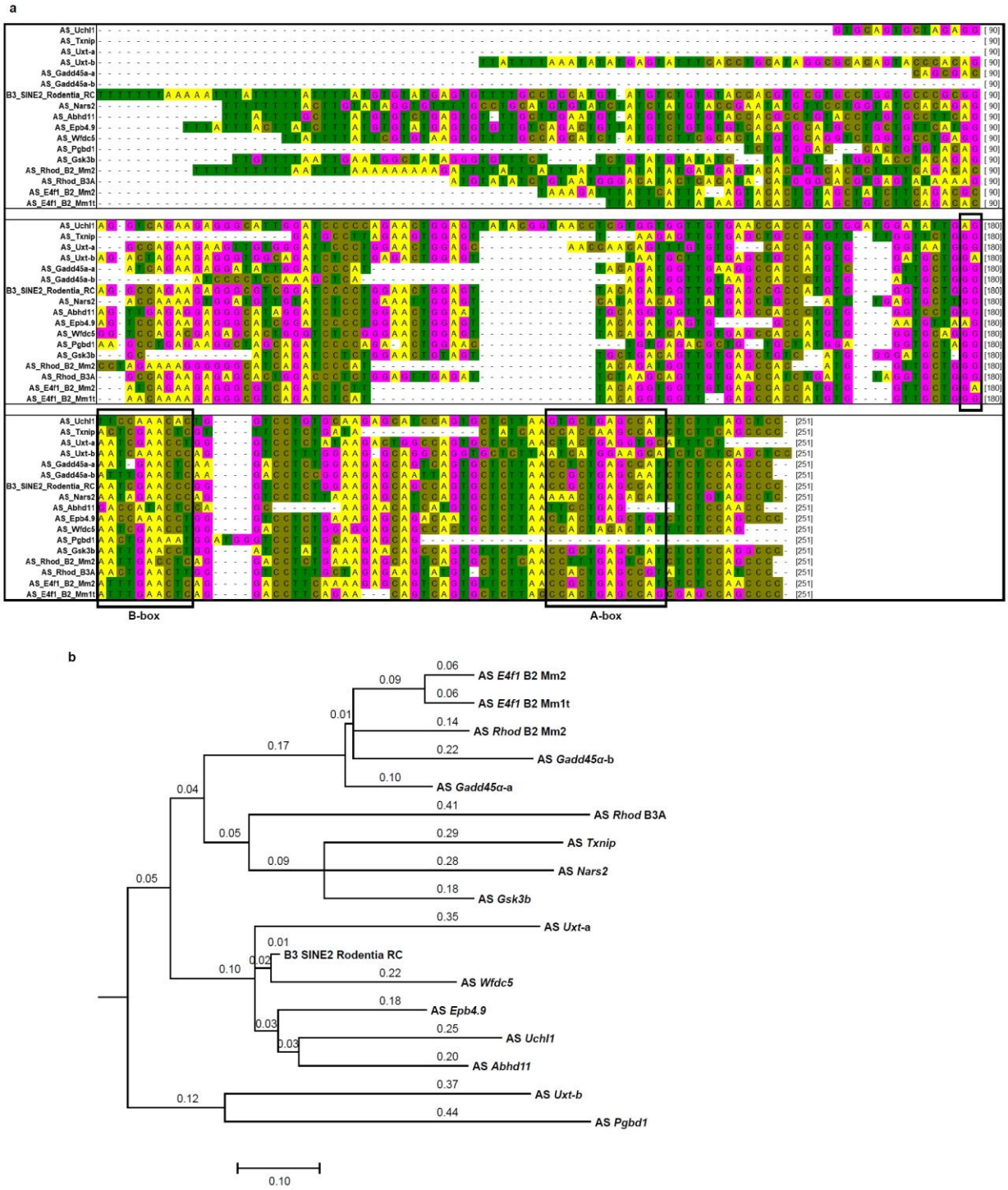
1365

1366

1367

1368

Extended Data Figure 3

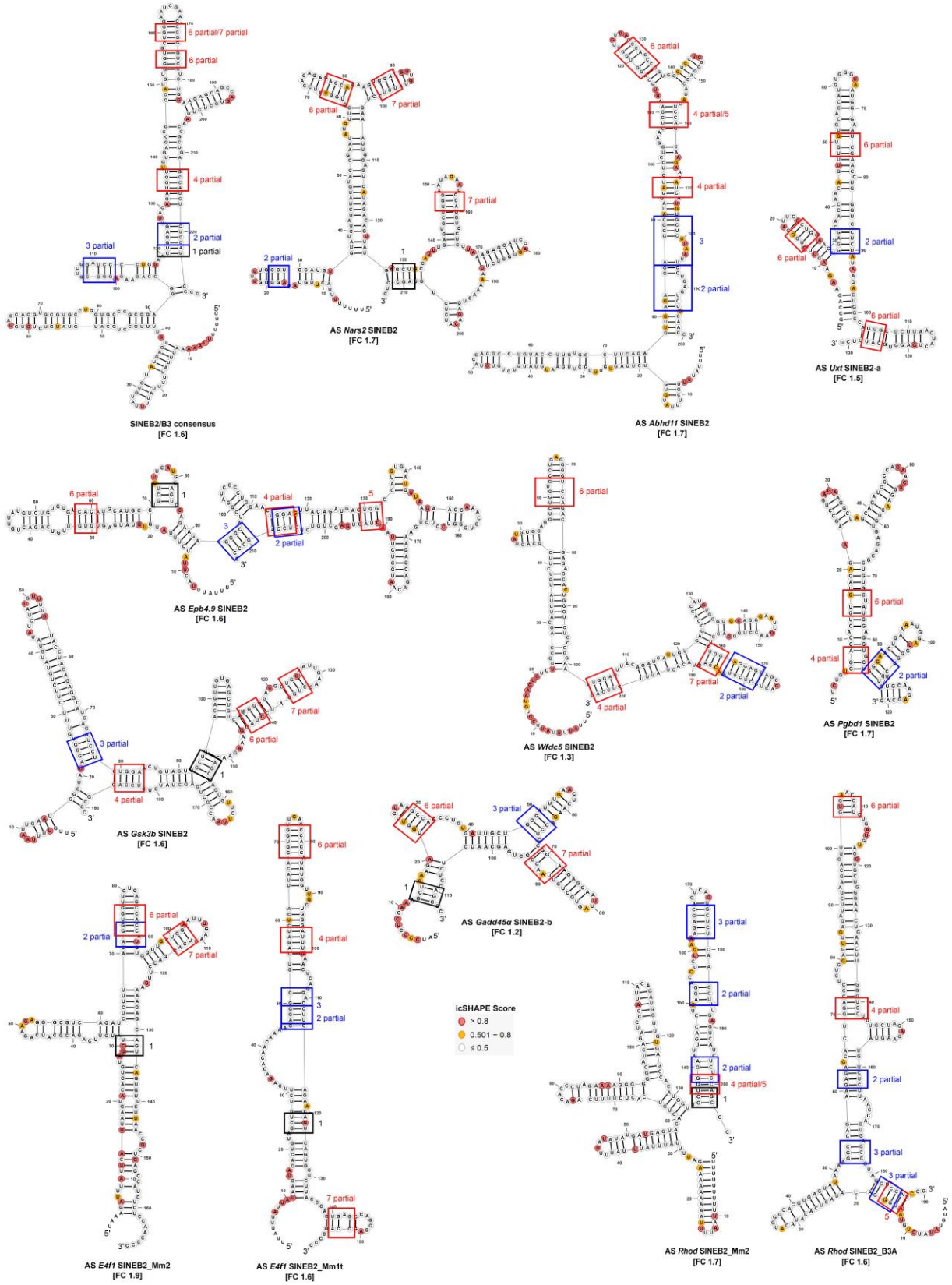


1369

1370 **Extended Data Figure 3. Multiple sequence alignment and phylogenetic analysis of SINEB2 RNA**
1371 **sequences.** (**a**) ClustalW sequence alignment of mouse SINEB2 RNA sequences used as effector domain
1372 in long and mini SINEUPs. Gaps in the alignment are indicated by -. Bases are highlighted in four different

colors. Black squares denote A and B boxes. Alignment is showing the sequence variation within members of the same family of SINE. (**b**) SINEB2 sequence alignment–based molecular phylogenetic analysis by the Maximum Likelihood method (Tamura 3-parameter model) to infer the evolutionary history. The tree with the highest log likelihood (−2415.7806) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. AS, antisense; B3 SINE2 Rodentia RC, the consensus sequence for SINEB2 subfamily B3 from the Repbase database.

Extended Data Figure 4



1402

1403 **Extended Data Figure 4. icSHAPE 2D structure models of mouse SINEB2 RNAs from natural antisense**
1404 **transcripts that are tested as effector domain in SINEUPs.** Regions that match the SINEUP motifs are
1405 marked by squares with their corresponding motif number. Blue square, AGG type motif; red square, UGG
1406 type motif. FC, GFP protein fold-change induced by the respective SINEB2. Nucleotide color indicates
1407 normalized icSHAPE reactivity score: red, ≥ 0.8; yellow, from above 0.5 to 0.8; white encircled with grey,
1408 ≤ 0.5; bases not encircled, no icSHAPE data available. AS, antisense.

1409

1410

1411

1412

1413

1414

1415

1416

1417

1418

1419

1420

1421

1422

1423

1424

1425

1426

1427

1428

1429

1430

1431

1432

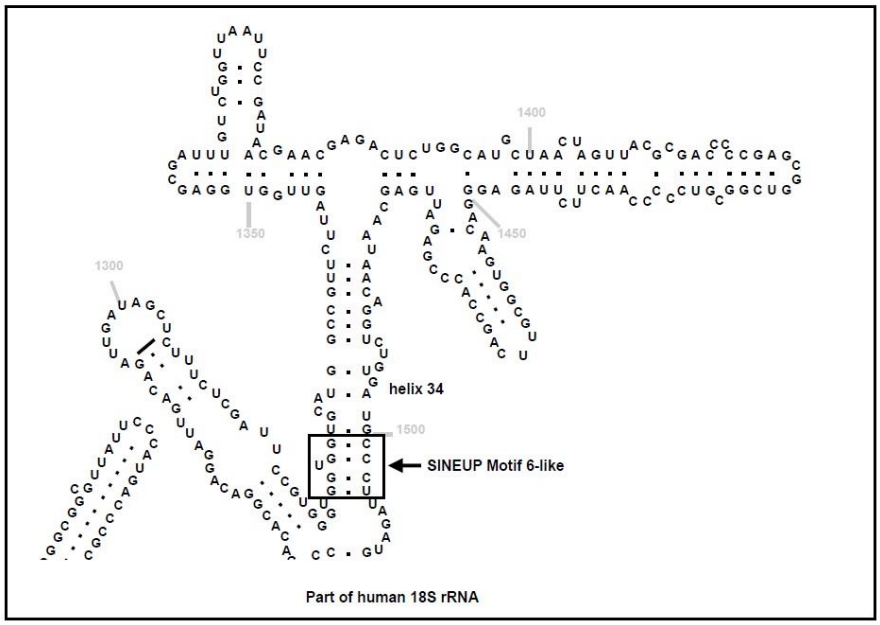
1433

Extended Data Figure 5

a

S. No.	SINEUP Motif	CoSSMos motif match	PDB ID	Positions	Molecule
1.	Motif 6	GGUGG UCCC	5A2Q	G1324-G1325-U1326-G1327-G1328 U1504-C1503-C1502-C1501	18S rRNA
2.	Motif 2	GGAGG CCCC	5JB3	G999-G1000-A1001-G1002-G1003 C1170-C1169-C1168-C1167	16S rRNA

b

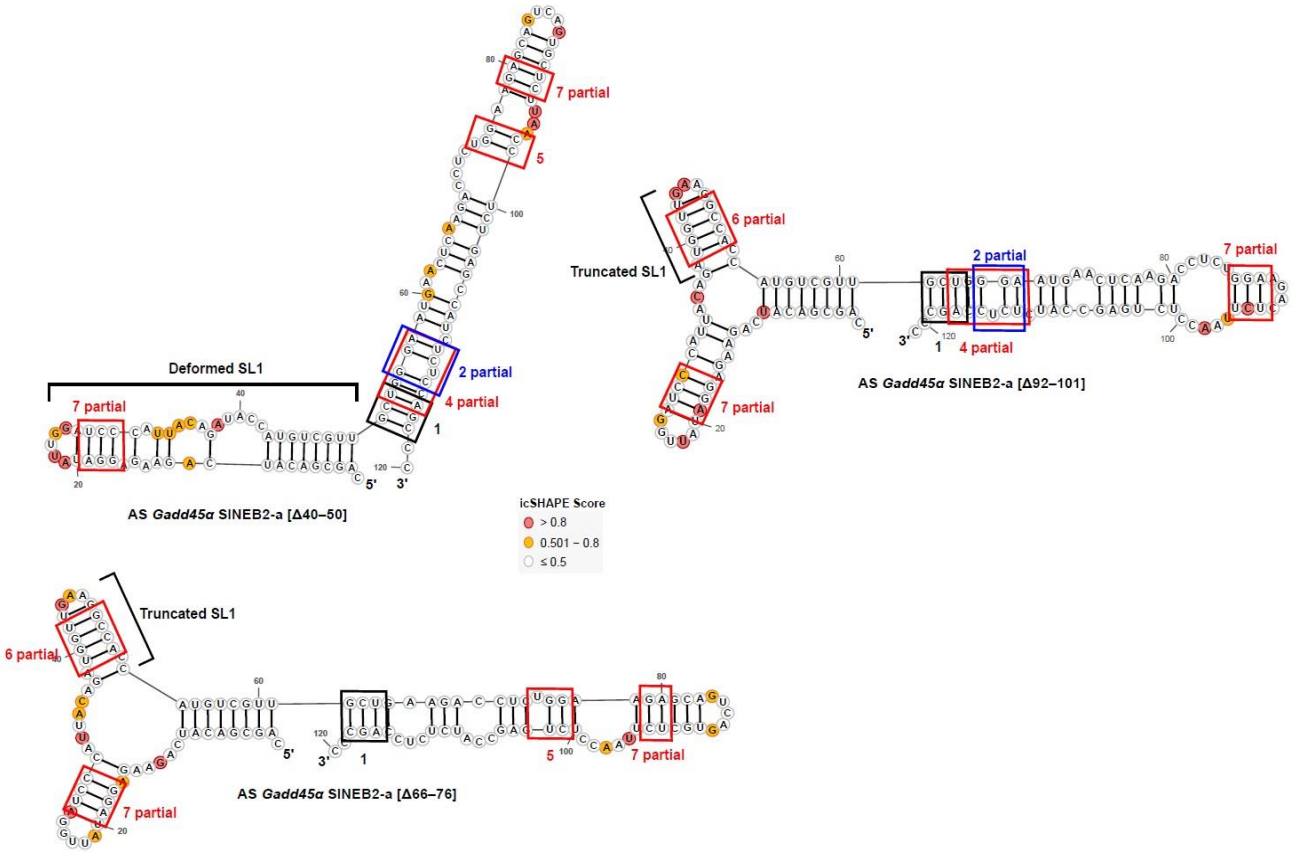


1434

1435 **Extended Data Figure 5. Similarity of SINEUP motifs to rRNA 3D structure motifs.** (**a**) Hits found in the
1436 Characterization of Secondary Structure Motifs database for 3D motifs that match the SINEUP structure
1437 motifs. (**b**) The region similar to SINEUP motif 6 in helix 34 of human 18S rRNA is framed in black and
1438 marked with an arrow.

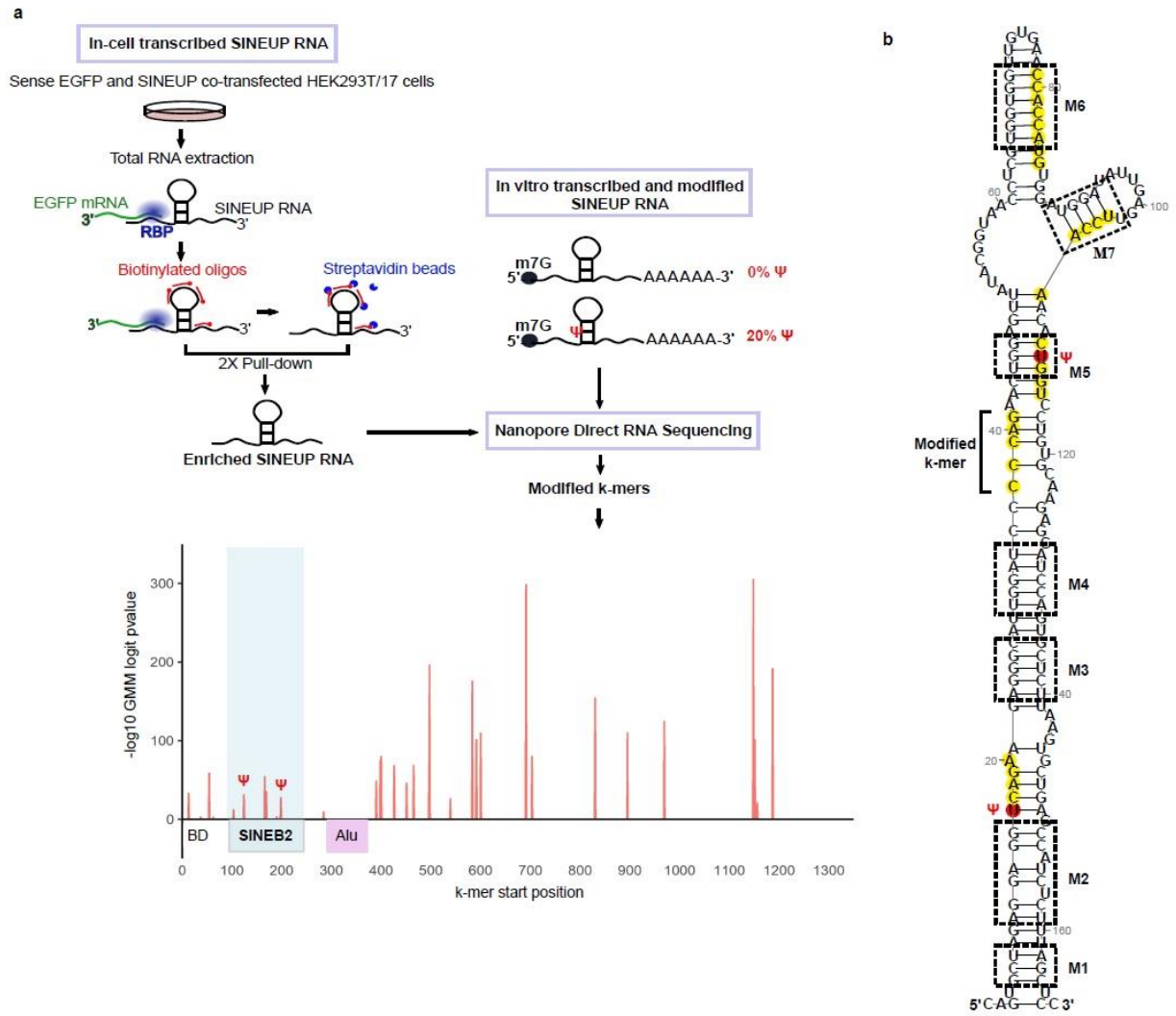
1439

Extended Data Figure 6

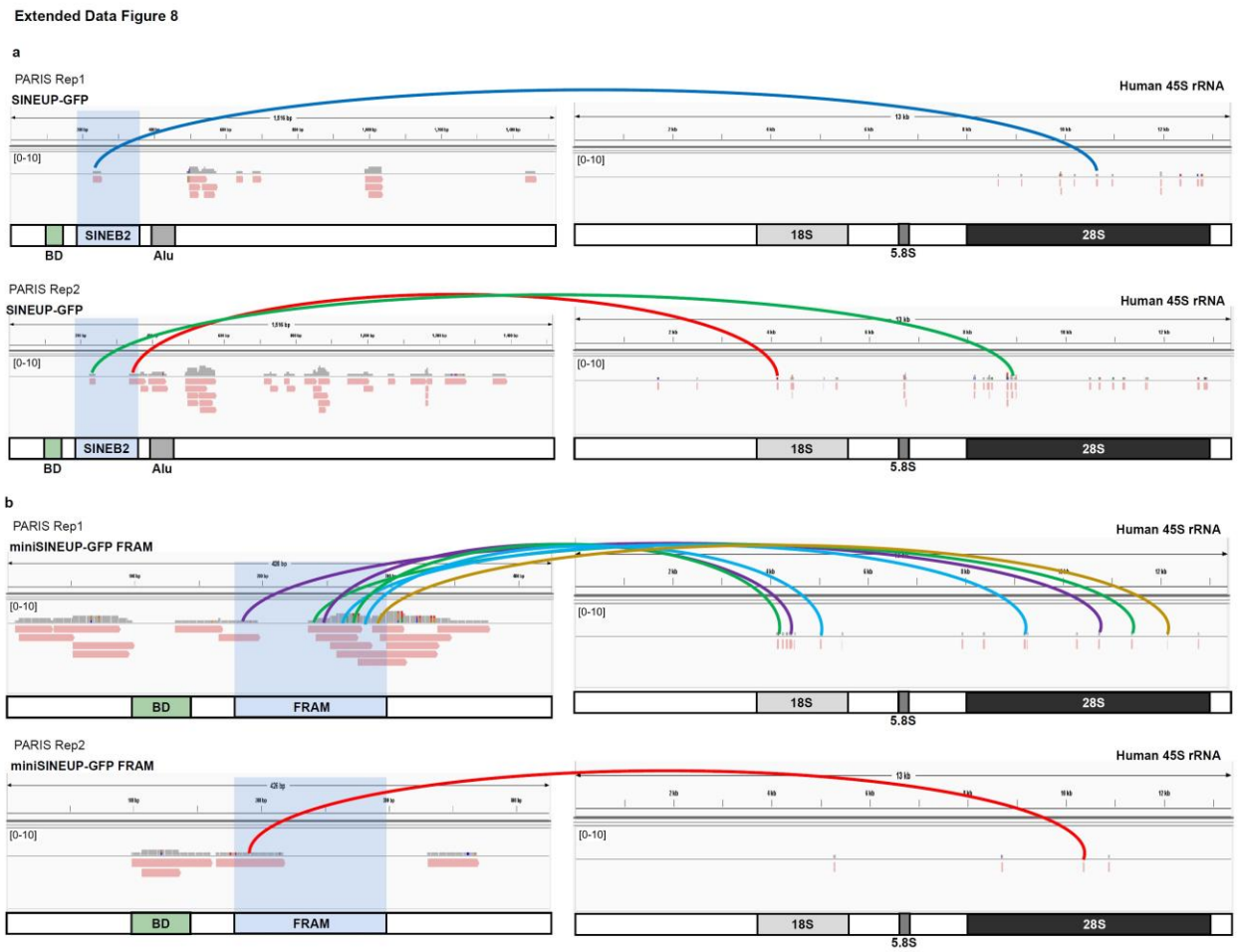


Extended Data Figure 6. SINEUP structure motifs in functionally inactive antisense (AS) *Gadd45α* SINEB2-a deletion mutants. icSHAPE derived structures of AS *Gadd45α* SINEB2-a deletion mutants [Δ40–50], [Δ66–76], and [Δ92–101]. Regions that match conserved SINEUP structure motifs are marked by different colors and numbers. Δ = deletion. Nucleotide color indicates normalized icSHAPE reactivity score: red, ≥ 0.8; yellow, from above 0.5 to 0.8; white encircled with grey, ≤ 0.5; bases not encircled, no icSHAPE data available. AS, antisense.

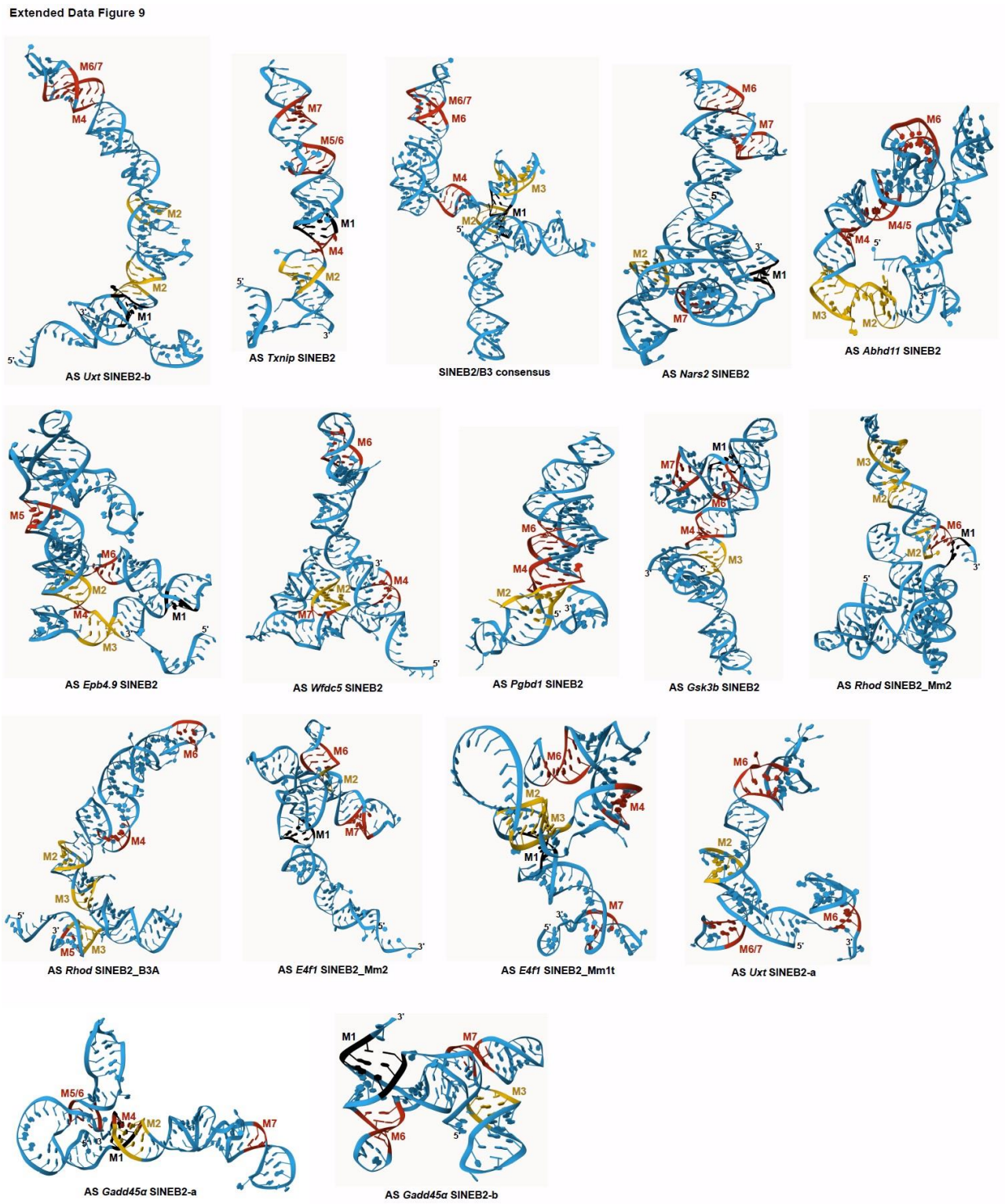
Extended Data Figure 7



Extended Data Figure 7. Discovery of modified bases in SINEB2 RNA. (**a**) An overview of SINEUP RNA modification analysis. SINEUP-GFP RNA (containing antisense Uchl1 SINEB2) transcribed in vitro was modified with 20% pseudouridine (Ψ), and a non-modified (0% Ψ) sample was used as control. In parallel, SINEUP-GFP and sense EGFP plasmids were co-transfected in HEK293T/17 cells, then SINEUP RNA that was transcribed in-cell was captured and enriched using specific biotinylated oligo probes and pull-down (twice) on streptavidin-coated magnetic beads. Purified samples of SINEUP-GFP RNA transcribed in-cell or transcribed and modified in vitro were sequenced by Nanopore direct RNA sequencing. Profiles of non-modified (0% Ψ) and modified SINEUP-GFP RNA transcribed in vitro were compared with that transcribed in-cell, and k-mers corresponding to the modified positions were identified in the in-cell transcripts. The red peaks in the graph mark the position of modified k-mers on the SINEUP transcript. The SINEB2 region is shaded in blue, and verified Ψ sites are marked. BD, binding domain. (**b**) Modified k-mers overlaid on 2D structure of antisense *Uchl1* SINEB2. k-mers are highlighted in yellow, Ψ sites are in red. SINEUP motifs are marked in black dashed squares with motifs 1–7 labelled as M1–M7.

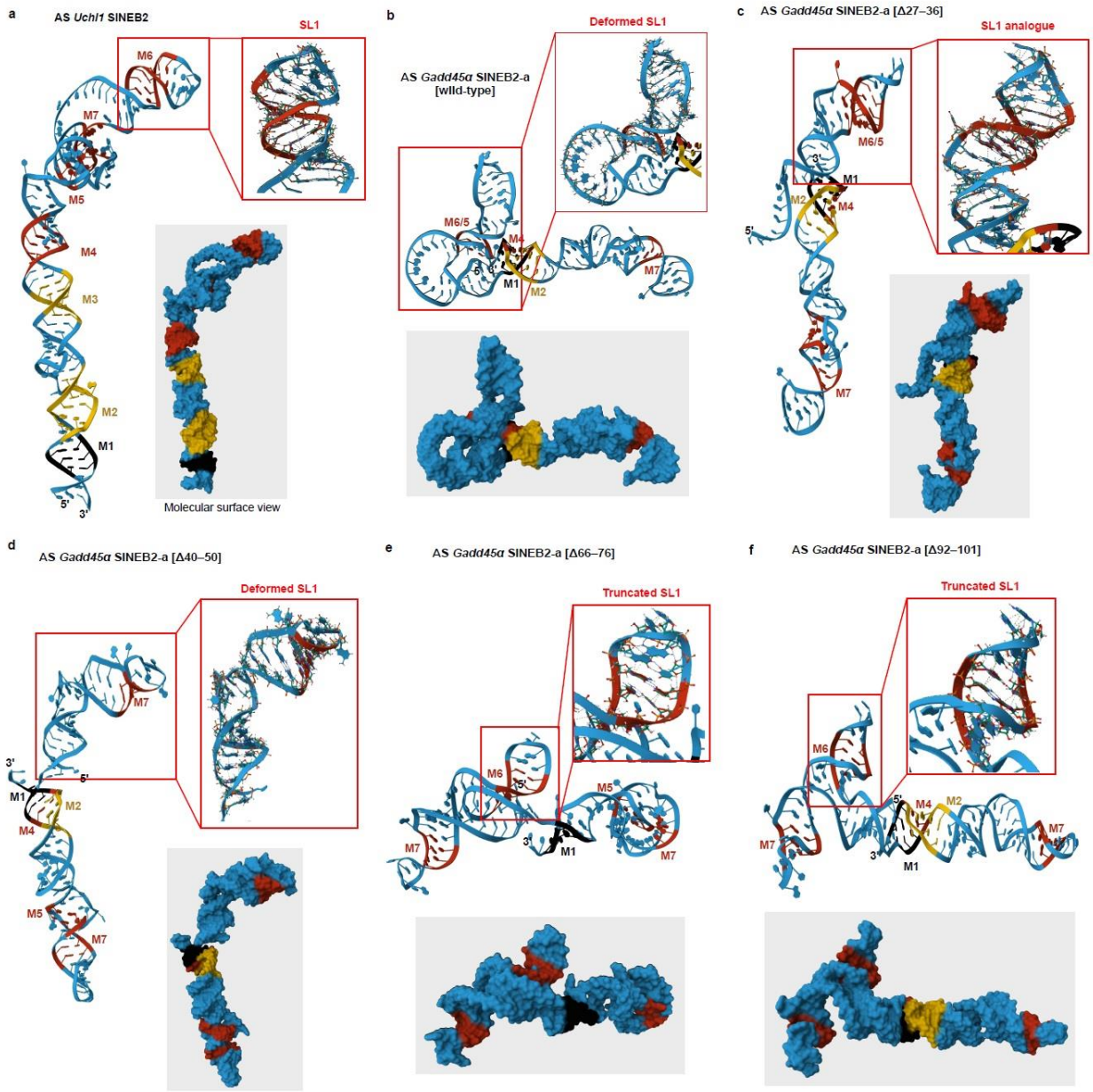


Extended Data Figure 8. SINEUP-rRNA duplex reads in PARIS sequencing data. (**a**) SINEUP-GFP (of antisense *Uchl1* SINEB2) RNA duplex reads with 18S, 5.8S, and 28S rRNA. Two biological replicates of the PARIS experiment are shown as separate tracks. SINEB2-rRNA interactions shown in Fig. 4 are indicated by respective colored arcs. (**b**) miniSINEUP-GFP FRAM RNA-rRNA duplex reads. PARIS rep1 and rep 2 are two biological replicates. The FRAM RNA-rRNA interactions displayed in Fig. 5 are marked by arcs in their respective colors.



Extended Data Figure 9. SINE RNA predicted 3D structure models based on in-cell 2D structures of mouse SINEB2 RNAs. SINEUP structure motifs are denoted from M1 to M7. AS: antisense.

Extended Data Figure 10



Extended Data Figure 10. Comparison of icSHAPE data-driven predicted 3D structure models of wild-type AS *Gadd45α* SINEB2-a and its deletion mutants with functional AS *Uchl1* SINEB2. (**a**) AS *Uchl1* SINEB2 3D model predicted from whole cell icSHAPE data. Predicted 3D models with enlarged view of the corresponding SL1 regions for AS *Gadd45α* SINEB2-a (**b**) wild-type (functionally inactive), its (c) functionally active deletion mutant [Δ27–36], and non-functional mutants (d) [Δ40–50], (e) [Δ66–76], and (f) [Δ92–101]. SINEUP structure motifs M1–M7 are marked. Red frame shows zoomed-in view of the region corresponding to SL1. The molecular surface views of the 3D models are shaded in gray. Δ = deletion.

1498 **Supplementary Table 1:** SINEB2 sequences tested as effector domain (ED) in (**a**) long SINEUP and (**b**) in
1499 miniSINEUP-GFP.

1500

1501 **Supplementary Table 2:** List of base-stacking positions in SINE RNA 3D models.

1502

1503 **Supplementary Table 3:** List of non-canonical interactions identified in SINE RNA 3D models.

1504

1505 **Supplementary Table 4:** Intermolecular interactions of SINEUPs with RNAs other than rRNA in PARIS2
1506 data.

1507

1508 **Supplementary Table 5:** (**a**) Primers used for QuikChange II site-directed mutagenesis. (**b**) Primers used
1509 for PCR-cloning of miniSINEUP-GFP constructs.

1510

1511 **Supplementary Table 6:** Oligos used in (**a**) icSHAPE, (**b**) seCLIP, and (**c**) PARIS library preparation.

1512

1513

1514

1515