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

38 for RNA–DNA, RNA–RNA, and RNA–protein regulatory complexes[5,6]. The combination of these different
39 domains provides the lncRNAs with their unique functional specificity and diversity[5]. Such "modular RNA
40 code" was validated in recent studies where short structure motifs were found to control the lncRNA
41 stability and function, such as 3′ end triple helix in MALAT1 and NEAT1[7,8], 5′ asymmetric G-rich internal
42 loop (AGIL) motif in mouse Braveheart (Bvht)[4], pseudoknots in human MEG3[9], and a terminal hairpin
43 stem-loop (SL1) in mouse lncRNA antisense to *Uchl1* mRNA[10-12]. In the antisense to *Uchl1*, the crucial SL1
44 region is within an embedded short interspersed nuclear element (SINE) from the B2 family, which
45 forms the effector domain (ED) that is essential for targeted mRNA translation upregulation[10,11].

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

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

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

76 LncRNA structures are known to act as a molecular scaffold for proteins[36,37], and in fact the binding of
77 heterogeneous nuclear ribonucleoprotein K (HNRNPK) with SINEUP and target mRNA complex is vital for
78 SINEUP-mediated upregulation of target mRNA translation[38]. HNRNPK aids in the export of SINEUP
79 complexes to the cytoplasm and co-sediments with SINEUP in free 40S ribosomal subunits and
80 monosome fractions[38], thus exhibiting two discrete functions in the nucleus and the cytoplasm. This
81 knowledge makes it imperative to examine how specific SINE RNA structures in different cellular
82 compartments interact with RNA binding proteins (RBPs) and support their distinct functions. In addition
83 to intramolecular structure, intermolecular interactions with other RNAs play a critical role in gene
84 regulation by non-coding RNA[39,40], notably the binding to rRNAs by linear and circular internal ribosome
85 entry sites (IRESs) for promotion of translation[41,42]. In fact, our previous molecular dynamics simulation
86 study on NMR solution structure of SINEB2 SL1 identified 3D motifs similar to those in rRNAs[43]. This data
87 and the association of SINEUPs with ribosomal subunits in polysome profiles[38] made us ponder the
88 possibility that direct interaction between SINE-derived RNAs and rRNAs may be a mechanism by which
89 SINEUPs enhance mRNA translation.

90 To solve the puzzle of SINE-transcribed RNA structure, we utilized a method called in vivo click selective
91 2′-hydroxyl acylation and profiling experiment (icSHAPE), which chemically probes flexible RNA bases
92 inside living cells and thus provides an advantage over other chemical or enzymatic methods[44]. We
93 present two-dimensional (2D) secondary structure models for 17 mouse SINEB2s-derived RNAs and a
94 human SINE FRAM RNA, providing for the first time experimentally verified in-cell 2D-structure models
95 of evolutionarily close and distant SINE-transcribed RNAs at this scale. We found that structural deletion
96 leading to the gain of an apical stem-loop (known as SL1) restores the activity in a SINEUP that is
97 functionally dormant (i.e., lacking the ability to significantly enhance the target mRNA translation).
98 Based on the 2D models, we discovered seven short conserved-structure motifs among the SINE RNAs
99 tested, three of which harbor modified RNA bases. The icSHAPE-guided structures for the nucleus and

100 cytoplasm showed that SINE RNAs in both mouse (SINEB2, derived from tRNA) and human (FRAM,
101 derived from 7SL RNA) rearrange into a cloverleaf-like structure in the cytoplasm and have partially
102 conserved structure motifs. We also observed that compartment-specific structures of SINE RNA are
103 associated with distinctive binding patterns of HNRNPK in the nucleus and the cytoplasm by using a
104 single-end enhanced crosslinking and immunoprecipitation assay (seCLIP)[45]. Additionally, when motifs
105 were compared in silico with those in the Protein Data Bank (PDB), we detected a remarkable similarity
106 of SINEUP 2D structure motifs with rRNA structure motifs. We also checked RNA–RNA chimera by the
107 psoralen analysis of RNA interactions and structures (PARIS2) method[40] and identified interactions of
108 mouse and human SINE-derived RNAs with human 18S and 28S rRNAs in the conserved structure motif
109 regions, which shed some light on SINEUP's mechanism of action. Finally, we used our experimental 2D
110 structure data to guide the prediction of structure models for 18 SINE-transcribed RNAs in three-
111 dimensions (3D) which can provide a base for future experimental 3D structure studies. In brief, this
112 study demonstrates that despite sequence variability among SINE-transcribed RNAs, short stretches of
113 conserved sequence and structure preserve the regulatory functions by providing binding sites for RBPs
114 and rRNA. These short motifs are likely to be stabilized by modifications and non-canonical base
115 interactions and are evolutionarily favored to maintain the SINEUP functionality even in distantly related
116 SINE-derived RNAs.

117

118 **Results**

119 **The SINEUP effect is a prevalent functional feature of mouse SINEB2s-derived RNAs**

120 To experimentally verify the translation-regulatory function of various mouse SINE-transcribed RNAs in
121 the SINEUP effector domain (ED), we used a previously established synthetic SINEUP-GFP model that
122 enhances EGFP mRNA translation (**Fig. 1a**). We swapped the SINEB2 from antisense *Uchl1* in SINEUP-GFP
123 with those from other mouse antisense to genes: *Txnip*, *Gadd45α*, and *Uxt* (**Supplementary Table 1a**).
124 RepeatMasker revealed the presence of an inverted SINEB2 repeat spanning from 1389 to 1495
125 nucleotide positions (107 bp long) in antisense *Txnip*, and both antisense *Gadd45α* and antisense *Uxt*
126 lncRNAs contain two inverted SINEB2 repeats in the region not complementary to the sense mRNA. Both
127 antisense *Uxt* SINEB2 sequences are from the B3 sub-family; in the FANTOM3 cDNA clone, one sequence
128 covered positions 160 to 290 (131 bp) and the other 774 to 960 (187 bp) (**Supplementary Table 1a**).
129 Based on their natural order of occurrence from the 5′ end of the transcript, the sequences were

130 named antisense *Uxt* SINEB2-a (131 bp, proximal to the 5′ end) and antisense *Uxt* SINEB2-b (187 bp,
131 distal to the 5′ end) (**Supplementary Table 1a**, **Fig. 1b**). In contrast, both SINEB2 repeats of antisense
132 *Gadd45α* belong to the B2_Mm2 sub-family (according to RepeatMasker nomenclature). We named
133 these two sequences antisense *Gadd45α* SINEB2-a (131 bp, proximal to the 5′ end) and antisense
134 *Gadd45α* SINEB2-b (114 bp, distal to the 5′ end) (**Supplementary Table 1a**). To understand whether
135 both SINEB2 sequences are essential for SINEUP function and how their order affects SINEUP activity,
136 we designed antisense *Gadd45α* and antisense *Uxt* constructs with a or b sequences and also with both
137 sequences in the natural (ab) or the inverted (ba) order (**Fig. 1b**). The shortest SINEB2 of those tested,
138 antisense *Txnip* SINEB2 (107 bp), enhanced the GFP protein level (**Fig. 1c** and **d**, **Supplementary Table 1**).
139 None of the four constructs with antisense *Gadd45α* SINEB2s could exert the SINEUP effect, regardless
140 of the order of SINEB2 a and b (**Fig. 1c** and **d**). Indeed, natural antisense *Gadd45α* was found to inhibit
141 translation of sense *Gadd45α* (**Extended Data Fig. 1**), which implies that the SINEUP activity of antisense
142 *Gadd45α* SINEB2 RNA is dormant, possibly due to the lack of functional regions. Interestingly, of the
143 antisense *Uxt* SINEB2 repeats, only SINEB2-b upregulated the GFP protein level significantly, and when
144 the position of the two SINEB2 sequences was switched from the natural order (ab switched to ba), their
145 SINEUP effect was obliterated (**Fig. 1c** and **d**). Upholding the post-transcriptional regulatory nature of
146 SINEUPs, none of the SINEUPs changed the EGFP mRNA expression level significantly (**Extended Data**
147 **Fig. 2a**).

148 To further expand analysis of the SINEB2-derived RNAs present in mouse natural antisense transcripts
149 (**Supplementary Table 1**), we used the shortest functional model of SINEUP-GFP, named miniSINEUP-
150 GFP[29]. SINEB2 sequences from antisense *Uchl1* and antisense *Uxt* enhanced GFP protein expression by
151 around 1.8–2.2-fold (**Fig. 1e**, Lanes 2 and 3 in the left graph). We additionally found that combining the
152 SINEB2 from antisense *Uchl1* with antisense *Uxt* SINEB2-b upregulated GFP protein production;
153 however, the protein fold change was lower than the effect of either SINEB2 independently (**Fig. 1e**,
154 Lane 4 in the left graph). We additionally tested three repeats of antisense *Uchl1* SINEB2 but the SINEUP
155 function remained similar to that with only one repeat, implying that these repeats do not function in a
156 synergistic or additive manner (**Fig. 1e**, Lane 5 in the left graph).

157 Since the SINEB2 sub-family B3 acted as a functionally effective SINEUP ED, we decided to check the
158 SINEUP potential of consensus subfamily B3 sequence from the Repbase database[27,46]. Indeed, the
159 sequence is functionally active against GFP (~1.6-fold change) (**Fig. 1e**, Lane 6 in the left graph), but not
160 to the level of antisense *Uchl1* SINEB2 RNA, which is a similar result to tests on other SINEB2 sequences

161 in the B3 sub-family from antisense lncRNAs to genes: *Nars2*, *Abhd11*, *Epb4.9*, *Wfdc5*, *Pgbd1*, and *Gsk3b*
162 (**Fig. 1e**, Lanes 7–10 in the left graph and Lanes 3–4 in the right graph).

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

173

174 **Functional SINEs form helical structures in the cell and consist of conserved short structure motifs**

175 To check any common sequence features between different SINEB2-transcribed RNAs showing the
176 SINEUP effect and to identify any functionally important regions, we performed multiple sequence
177 alignment using ClustalW (**Extended Data Fig. 3a** and **b**). This revealed conserved regions of RNA pol III
178 internal promoters known as A box and B box[47] (**Extended Data Fig. 3a**) along with sequence diversity
179 between SINEB2 elements tested in this study. Moreover, in the phylogenetic tree based on the multiple
180 sequence alignment, the functionally active SINEB2s that significantly upregulated GFP translation were
181 not clustered separately from the inactive SINEB2s that failed to do so (**Extended Data Fig. 3b**). In other
182 words, on the basis of sequence, the functional SINEB2 RNAs do not possess any unique features that
183 confer a functional advantage over the inactive SINEB2 RNAs. This lack of conserved sequence domains
184 led us to examine the in-cell secondary structure of SINEB2s by using the icSHAPE method to map the
185 RNA secondary structure inside living cells[44,48].

186 We compared the icSHAPE antisense *Uchl1* SINEB2 RNA secondary structure with our previous in vitro
187 chemically probed structure[11], which had identified four internal-loops (IL1–IL4) and three stem-loops
188 (SL1–SL3) (**Fig. 2a**). Both methods gave a notably similar helical structure for antisense *Uchl1* SINEB2
189 RNA (**Fig. 2a**). The icSHAPE-based structure comprises three internal loops similar to the IL2, IL3, and IL4
190 found in the chemical footprinting-based structure, with slight differences in the sequence composition.

191 The in-cell structure lacks SL3, which indicates the influence of intracellular factors and dynamic
192 structural changes compared to the less complex conditions in vitro. Interestingly, SL1 and SL2 are
193 conserved between the structures derived by these two different methods, which suggests that they
194 represent the most dominant and stable structure components in the ensemble. The structural
195 conservation is also understandable knowing that SL1 is crucial for SINEUP function[11]. In a similar way,
196 secondary structure modelling was performed for other SINEB2s-derived RNAs (**Fig. 2b** and **c**, **Extended**
197 **Data Fig. 4**) and revealed that they all retained helical features with a terminal GC-rich stem-loop similar
198 to SL1 of antisense *Uchl1* SINEB2, even though all SINEB2 RNAs tested here had diverse global structures
199 with variable sequence length and composition.

200 The importance of short structure motifs like SL1 in the SINEUP function prompted us to look for other
201 local motifs that are shared between different SINEB2-transcribed RNAs. For this purpose, we utilized
202 the ExpaRNA tool[49] and, using the antisense *Uchl1* SINEB2 RNA structure as a base, searched for its exact
203 sequence and structure patterns in other icSHAPE-derived SINEB2 RNA structures. From this search, we
204 discovered seven GC-rich structure motifs partially conserved among the SINEB2 RNA structures tested
205 in this study (**Fig. 2a–c**, **Extended Data Fig. 4**). Motifs 2 and 3 are comprised of 'AGG' trinucleotide(s),
206 while motifs 4, 5, 6, and 7 are 'UGG' enriched. Motif 6 lies within the SL1 region making it a potentially
207 functional motif. Furthermore, we wondered if the motifs are conserved in SINE-transcribed RNAs that
208 are evolutionarily distant from mouse SINEB2 elements, such as human SINE FRAM, which has been
209 proven to be effective as a SINEUP against both endogenous and exogenous mRNA targets[31]. Indeed,
210 the structure motifs in the whole-cell icSHAPE structure of FRAM RNA provided five matching regions
211 (**Fig. 2d**). This hints that conservation of such short motifs may contribute in functional preservation.

212

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

214 Considering the conservation of short local structure motifs in functional SINE-transcribed RNAs, we
215 theorized that dormant SINEUPs, here the SINEB2s that failed to significantly upregulate GFP protein
216 expression in **Fig. 1c**, might include embedded inhibitory regions that prevent their proper folding into
217 functional domains. To test this, we chose a functionally inactive SINEB2 sequence, antisense *GADD45α*
218 SINEB2-a, and randomly deleted 10–12-nucleotide stretches to change its structure, to assess if
219 mutations can restore SINEUP activity and transform a dormant SINEUP into an active one. In total, we
220 made five deletion mutants at positions 27–36, 40–50, 66–76, 92–101, and 109–120 from the 5′ end of

221 SINEB2 and co-transfected the mutants with sense EGFP into HEK293T/17 cells, as done in the previous
222 SINEUP-GFP functional analysis experiment. Western blot data revealed that out of the five mutants
223 Δ27–36 could significantly upregulate GFP protein level compared to full-length antisense *GADD45α*
224 SINEB2-a (**Fig. 2e**). We compared the predicted RNA structures of these deletion mutants with the
225 experimentally verified full-length structures of wild-type SINEB2 and that of antisense *Uchl1* SINEB2
226 (**Fig. 2a**, **f**, **Extended Data Fig. 5**) and discovered that deleting positions 27–36 from the wild-type
227 resulted in the gain of an SL1-like structure, which in turn activated the SINEUP function (**Fig. 2f**). Out of
228 all the mutants, Δ27–36 shared the most similarity in global structure and arrangement of SINEUP motifs
229 with functional antisense *Uchl1* SINEB2, i.e., an elongated helical structure with two stem loops, where
230 SL1 is taking the apex position (**Fig. 2a**, **f**). This suggested that some structural features in the wild-type
231 antisense *GADD45α* SINEB2-a are detrimental to function, whereas the acquisition of short local
232 structure motifs leads to functionally favorable RNA folding, which can restore SINEUP function. The
233 other functionally inactive mutants were also predicted to contain some partial matches to the structure
234 motifs but with different global structures and varying placement of the motifs along the structure
235 (**Extended Data Fig. 5**).

236

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

238 Our SINEUP icSHAPE data from whole cell lysate showed a set of common motifs among different
239 SINEB2s, yet it was an average of whole-cell structure ensemble and did not help to infer structure
240 dynamics within the cell. RNA structure is closely related to cellular processes and can acquire distinct
241 conformations depending on the cellular compartment[32]. Understanding the different RNA structures
242 becomes more crucial in the case of SINEUPs because they are involved in post-transcriptional
243 translation regulation, a function specific to cytoplasm. Therefore, it is logical to examine the structure
244 dynamics of functional SINEB2s-derived RNAs both in the nucleus and the cytoplasm, to better
245 understand the contribution of structural conformations to the biological processes specific to these
246 subcellular regions. To achieve this, we made icSHAPE libraries from nucleoplasmic and cytoplasmic
247 fractions of cells transfected with SINEUP-GFP carrying either mouse antisense *Uchl1* SINEB2 or human
248 SINE RNA FRAM. The 2D structure models based on this icSHAPE data revealed that the base of the
249 SINEB2 RNA structure is mostly stable between the nucleus and the cytoplasm, with only a slight
250 variation observed around the region of motif 2, being slightly more open in the nucleus than in the
251 cytoplasm (**Fig. 3a**, **b**), and a similar structure difference in human FRAM-transcribed nuclear RNA (**Fig.**

252 **3c**, **d**). In addition, the internal loop between motifs 5, 6, and 7 in the nuclear mouse SINEB2 RNA goes
253 through a structural rearrangement in the cytoplasm and transforms into a cloverleaf-like structure (**Fig.**
254 **3a**, **b**). In the case of human FRAM RNA, both nuclear and cytoplasmic structures are quite similar and
255 form a cloverleaf-like structure, though the cytoplasmic structure is slightly more structured (**Fig. 3c**, **d**).
256 This could be possibly due to the transient nature of any alternative structures, as icSHAPE structure
257 represent only the average of the structure ensemble.

258 The nucleus-specific internal loop structure might play a role in the export of SINEUP RNA to the
259 cytoplasm, whereas the cytoplasmic cloverleaf-like structure with SL1 might be crucial for interaction
260 with translation machinery. Interaction with HNRNPK is crucial in both of these SINEUP processes[38]. To
261 check how the different nuclear and cytoplasmic structures of SINEB2 RNA are associated with HNRNPK,
262 we prepared seCLIP[45] libraries from the nuclear and cytoplasmic fractions of SINEUP-GFP co-transfected
263 cells (**Fig. 3e**). The results showed that in the nucleus HNRNPK interacts with short stretches of SINEB2
264 coinciding in motifs 1, 3, and 4 (**Fig. 3b**), but in the cytoplasm HNRNPK occupied a long double-stranded
265 region covering the 5′ end of the stem in all the motifs except motif 7. Such disparate interactions could
266 possibly arise from distinct functions of HNRNPK in SINEUP RNA transport and in translation that are
267 specific to the cellular compartment.

268

269 **SINEUP structure motifs interact with rRNA**

270 We wanted to further investigate the role of SINEB2 RNA structure in translation because it has features
271 reminiscent of tRNA; it forms a cloverleaf-like structure in the cytoplasm and possesses two
272 pseudouridine (Ψ) sites at 16 bp and 112 bp (**Extended Data Fig. 6**). CircRNA IRESs[42] and many other
273 IRESs are known to interact with 18S rRNA[41,50] to induce cap-independent translation, however the
274 region by which SINEUP-RNA interacts with ribosomal subunits and polysomes is not known. Since our
275 observations highlighted the importance of SINEUP motifs, we checked the PDB database for 3D RNA
276 structure motifs that were akin to our SINEUP structure motifs. To achieve this, we employed a
277 sequence- and structure-specific search for each of the seven SINEUP motifs in the RNA Characterization
278 of Secondary Structure Motifs (CoSSMos) database[51]. This resulted in hits that matched 16S rRNA (PDB
279 ID: 5JB3) and 18S rRNA (PDB ID: 5A2Q) to 2D structures for SINEUP motifs 2 and 6, respectively
280 (**Extended Data Fig. 7a**). The present finding is interesting considering that matching motifs are very
281 similar to SINEUP motifs derived in-cell in terms of sequence and structure. On closer inspection, we

282 observed that the matching region to 18S rRNA is within helix 34, which forms the head of the mRNA
283 entry channel[52] (**Extended Data Fig. 7b**).

284 Considering the similarities in sequence and structure, we wondered whether such motifs in SINEB2
285 RNA can interact with rRNA. To get a better understanding of SINEUP–rRNA interactions, we employed
286 PARIS2, which has proved useful in capturing direct mRNA–rRNA short interactions[40,53]. We prepared
287 PARIS2 libraries from HEK293T/17 cells transiently expressing SINEUP-GFP (antisense *Uchl1* SINEB2) or
288 miniSINEUP-GFP (FRAM) along with their target sense EGFP mRNA and sequenced them on a next-
289 generation sequencing platform. However, it resulted in the capture of only a few incidents of SINEUP-
290 rRNA chimeras (**Fig. 4**, **Fig. 5**, **Extended Data Fig. 8**), which is understandable given the technical
291 limitation of psoralen crosslinking and inefficient detection of scarcely expressed RNA duplexes and
292 highly dynamic interactions in PARIS2[40,54]. Nevertheless, it was intriguing to see interactions of mouse
293 SINEB2 element with human rRNA in both small and large subunits of ribosome (**Fig. 4a**, **b**, and **c**). We
294 discovered that the 3′ side of motifs 1 and 2 of SINEB2 RNA interacted with a region in helices 14 and 15
295 of 18S rRNA (Int 1 in **Fig. 4a** and **c**). On the other hand, the region between motifs 5 and 6 formed
296 contacts with domain III helices 55, 56, and 59 of 28S rRNA. Motif 5 and the flanking region also
297 interacted with domain I, helix 25, expansion segments ES7c and ES7d of 28S rRNA (Int 2 and 3 in **Fig. 4b**
298 and **c**).

299 Interestingly, functionally active human SINE-derived RNA FRAM displayed conservation of this rRNA
300 interaction trait. With better expression of FRAM RNA, we were able to detect more instances of FRAM-
301 rRNA chimera than for SINEB2 RNA. Similar to SINEB2, the 3′ side of motifs 1 and 2 and of motifs 5 and
302 6 of FRAM RNA formed contacts with helix 15 of 18S rRNA (Int 2 in **Fig. 5a**, **b**, **c**). In addition, the FRAM
303 RNA stem loop analogous to SL1 (containing motifs 2/3, 5, and 6) interacted with 18S rRNA helix 21,
304 es6b and es6c (Int 1 in **Fig. 5a**, **b**, **c**), and helices 37 and 38 made chimera with the 3′ end region of
305 FRAM RNA (Int 3 in **Fig. 5a**, **b**, **c**).

306 Moreover, the 5′ side of FRAM RNA flanking motif 1 interacted with 28S rRNA domain III helix 48 and
307 49a (Int1 in **Fig. 5d**, **e**, **f**), and helix 50 and 52 (Int5 **Fig. 5d**, **e**, **f**), whereas FRAM SL1 region interacted
308 with 28S rRNA domain IV, helix 63, ES27b and with domain I, helix 25, ES7c (Int 2 and 3 in **Fig. 5d**, **e**, **f**).
309 The short stem-loop at the 3′ end of FRAM structure was involved in interaction with 28S rRNA domain
310 V, helix 81 (Int4 in **Fig. 5d**, **e**, **f**).

311 Notwithstanding the shallow SINEUP PARIS data, the presence of SINEUP-rRNA chimeras in addition to
312 our knowledge of SINEUPs' fate in the polysomes[38], supports the notion of SINEUP–rRNA interaction.

313

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

315 The SINE 2D structure models helped us to understand the conservation and functional importance of
316 local short motifs in maintaining the SINEUP effect by interacting with RBP and rRNA. However, they
317 lacked information about non-canonical intramolecular interactions that stabilize the RNA structure and
318 affect the availability of interaction sites in 3D. This results in canonically unpaired regions in the 2D
319 structures that are actually non-canonically paired in 3D[55]. Therefore, we took advantage of in vivo SINE
320 RNA 2D structure information and incorporated it into the machine translation of 3D structure models.
321 Thus, we created 3D models for 17 mouse and one human SINE-derived RNAs (**Fig. 6** and **Extended Data**
322 **Fig. 9**). The 3D models revealed that the SINEUP motifs are stabilized by base-stacking interactions along
323 the helical structure and that bases in the internal loop and stem-loop regions form non-canonical
324 interactions adding more complexity to the structure (**Fig. 6 a**,**c**,**e**,**g** and **Supplementary Tables 3** and **4**).
325 The molecular surface view of the nuclear SINEB2 RNA show an elongated helical structure where SL1
326 takes the apical position (**Fig. 6b**). Interestingly, human nuclear FRAM RNA, which has a cloverleaf-like
327 conformation in 2D (**Fig. 3d**), folds similarly to nuclear SINEB2 RNA in 3D, with the three arms positioned
328 in such a way that all of the motifs are in the central arm, SL1 is at the apex, and the other two arms are
329 close to the central one (**Fig. 2d** and **Fig. 6 e**,**f**). In the cytoplasm, both SINEB2 and FRAM RNAs fold into a
330 cloverleaf-like conformation where SL1 moves from the center to the left position and in FRAM RNA the
331 other two arms open up (**Fig. 6 c**,**d**,**g**,**h**).

332

333 **Discussion**

334 There are about 175 families of SINE scattered across different domains of life[56]. Although generally
335 considered as parasite elements of the genome, a growing number of studies point to their gene-
336 regulatory activity, such as enhancer-like function, polyadenylation signal, transcription factor binding
337 sites, mRNA nuclear retention, and translation upregulation[10,15-18,23]. However, there are no extensive
338 studies on the sequence and structural features of functional SINE-transcribed RNA elements. Here, we
339 examined the SINEUP activity of 15 as-yet-uncharacterized SINEB2-derived RNAs from mouse natural
340 antisense transcripts and assayed their function in GFP protein upregulation. Most (12 out of 15) of the

341 SINEB2 RNAs showed SINEUP activity, but the remaining 3 (antisense *Gadd45α*-SINEB2-a and -b, and
342 antisense *Uxt* SINEB2-a) failed to upregulate GFP protein expression significantly under the current
343 experimental settings. The shortest SINEB2 RNA to enhance GFP protein expression, antisense *Txnip*
344 SINEB2, was found to be 107 nucleotides long, which is half the length of the longest functional natural
345 SINEB2-derived RNAs tested here (antisense *Nars2* SINEB2 and antisense *Epb4.9* SINEB2). This suggests
346 that the SINEUP class of RNAs in nature is vast and diverse and is likely to be a much larger class of
347 antisense RNAs than yet explored. Functional observation of the consensus SINEB2 sub-family B3
348 sequence also supported this, although it enhanced GFP translation to a lesser extent than the highly
349 functional antisense *Uchl1* SINEB2 RNA. This might be because consensus sequences represent the most
350 frequent occurrence of a particular nucleotide in a multiple sequence alignment of related sequences,
351 which may not necessarily be functionally optimal. We observed that in most cases only one repeat of
352 SINEB2 RNA as the effector domain is sufficient to exert the SINEUP effect, possibly because two
353 SINEB2-derived RNAs (which are generally GC-rich) may share some short complementary regions that
354 can interact and hinder the proper RNA folding into key functional features or may compete for binding
355 proteins and RNA partners. Another interpretation of this non-synergistic nature of these functional
356 SINEB2 RNAs is that SINEUP-induced protein upregulation is used to fine-tune translation in the cell,
357 keeping the protein within a physiological range of up to 2–3-fold change. Incidentally, this characteristic
358 makes SINEUPs ideal for therapeutic studies where proteins need to be regulated within a physiological
359 range. The sequence diversity within the same B3 sub-family or in comparison with other sub-families
360 such as B3A, B2-Mm2, and B2_Mm1t (as per Repbase nomenclature)[27] does not prevent the SINEUP
361 function. This observation confirms the universality of the SINEUP function in the SINEB2 repeat family
362 and the flexibility of its members to maintain key functional features despite varying length and
363 composition of RNA sequences. It would be interesting to explore other families of SINE elements in
364 other species and across a wide variety of mammalian and eukaryotic genomes in the future.

365 Noting the sequence diversity of functional SINEB2-derived RNAs and the importance of RNA structure
366 in non-coding RNA functions, we looked at the role of the SINE RNA secondary structure in the SINEUP
367 phenomena in living cells by using icSHAPE experiments. Consequently, we noticed a high correlation
368 between the in-cell 2D structure of antisense *Uchl1* SINEB2 RNA and the previously reported structure
369 from chemical footprinting[11]. Moreover, we found that the key functional SL1 structure was conserved
370 among various SINEB2s RNAs. When the functionally dormant antisense *Gadd45α* SINEB2-a RNA that
371 lacks an SL1 region was mutated to change its structure, one of the mutants (Δ27–36) acquired a
372 properly folded SL1 structure and the predicted 2D structure was similar to the icSHAPE structure of

373 antisense *Uchl1* SINEB2 RNA, resulting in SINEUP activity gain. This further emphasizes the functional
374 significance of this short structure in SINEUP-induced gene regulation.

375 We discovered 7 conserved structure motifs among the icSHAPE data-driven 2D structure models of all
376 the SINEB2-derived RNAs tested here, though only partial matches in sequence were observed, given
377 the sequence variation. Since we observed partial matches to these structure motifs in the functionally
378 inactive SINEs also, we hypothesize that the SINEUP function is influenced not just by the mere presence
379 of these structure motifs but also by their relative positions on the structure. Some structures might be
380 sterically more exposed to interaction with molecular machinery in the 3D space.

381 We recently reported that deletion of certain regions in antisense *Uchl1* SINEB2 RNA is detrimental to
382 SINEUP function[12]. Several of these deleted regions correspond to the SINEUP structure motifs identified
383 in this study: Δ100-111 disrupts motif 7, Δ130-141 is the deletion of the 3′ end side of motifs 3 and 4,
384 and Δ5-14 disrupts motifs 1 and 2[12]. Interestingly, their Δ65–74 mutant[12], which deletes part of motif 6,
385 does not show a significant decrease in the SINEUP function, whereas another study demonstrated that
386 deletion of SINEB2 RNA SL1 consisting of the entire motif 6 abolishes SINEUP activity[11]. The in-cell
387 structure of human SINE-derived FRAM RNA also showed a partial match for SINEUP motifs and the
388 presence of a structural feature analogous to SL1. These observations point at SINEUP structural motifs
389 that are critical for SINEUP function.

390 Mouse SINEB2s originate from tRNA element[56,57], and pseudouridylation of tRNA bases is reported to be
391 crucial for its stability and function[35]. Not only tRNAs, but all the RNAs involved in translation contain
392 several Ψ-modified sites[34,35]. This was also found to be the case for SINEUPs in this study. We identified
393 two Ψ modifications at positions 16 and 112 of antisense *Uchl1* SINEB2 RNA. The Ψ at position 112 is
394 inside motif 5, and that at position 16 is just at the motif 2 boundary. This indicates that
395 pseudouridylation of these SINEB2 RNA positions might help to stabilize the structure of these motifs by
396 forming extra hydrogen bonds by N1-H position of Ψ or by base-stacking, which is an intrinsic property
397 of Ψ[34,58].

398 We further analyzed the structural dynamics of both SINEB2 and FRAM RNAs in the nucleus and
399 cytoplasm and identified two regions that differ between their nuclear and cytoplasmic structures, with
400 motif 2 and the region between motifs 5 and 6 of SINEB2 RNA slightly more open in the nucleus than in
401 the cytoplasm. Less structured regions were also observed in the nuclear structure of FRAM RNA. We
402 wondered whether these nucleus-specific structures can provide binding sites for proteins involved in

403 nuclear transport. We knew that binding to HNRNPK is crucial for cytoplasmic co-localization of SINEUP-
404 GFP and sense mRNA complex and their association with polysomes[38]. Furthermore, HNRNPK is
405 reported to interact with c-myc IRES and stimulate its cap-independent translation[59]. Since HNRNPK
406 participates in both nuclear transport and translation regulation, we checked if structurally dynamic
407 regions have any effect on its interaction in subcellular compartments. Indeed, we observed distinctive
408 binding patterns of HNRNPK in the nucleus and cytoplasm; in the nucleus, it interacts with short regions
409 in motifs 1, 3, and 4 of SINEB2 RNA, but in the cytoplasm it interacts with a much longer double-
410 stranded region covering all but one motif. These particular binding patterns could be connected to
411 different functions of HNRNPK in the nucleus and cytoplasm.

412 In the cytoplasm, the loop region between motifs 5, 6, and 7 of mouse SINEB2 RNA assumes a
413 cloverleaf-like structure, formed from a multiloop where four stems converge. One stem is SL1, which
414 harbors motif 6, another stem consists of motif 7, and a disrupted motif 5 makes the top of the central
415 stem where the Ψ at position 112 is situated at the multiloop branching point. Interestingly, though the
416 nuclear and cytoplasmic structures of FRAM RNA are highly similar, we discerned a similar cloverleaf-like
417 structure. icSHAPE structures are generally enriched for the most stable and dominant structural
418 conformations in the cell, so any alternative short-lived structures might not have been captured. It is
419 fascinating to see that mouse SINEB2 RNA that originated from tRNA and human FRAM RNA that
420 originated from 7SL RNA[57] share common characteristics: short conserved structural motifs, a cloverleaf-
421 like structure in the cytoplasm, and positive regulation of translation.

422 In addition to the intramolecular structure, we were interested to better understand the intermolecular
423 interactions of SINE-derived RNAs with translation machinery. Interaction with 18S rRNA has been
424 reported to be crucial for both linear and circular IRESs[41,42,50]. In view of the knowledge mentioned
425 earlier that SINEUP-GFP RNA co-sediments with free small subunits of ribosomes and monosomes, we
426 checked the interaction of SINEB2 RNA with rRNAs. We found similarity between the structure of the
427 motif 6 sequence in 2D and that of a motif in 18S rRNA in 3D, and the structure of motif 2 was found to
428 be similar to a region in 16S rRNA in 3D. The matching region in 18S rRNA belongs to helix 34, which
429 forms the roof of the mRNA entry channel. The similarity of the SL1 region of SINEB2 RNA with rRNA was
430 previously noticed in a molecular dynamics simulation based on NMR data[43], and here in-cell analysis
431 again showed that the conserved and functionally crucial SINEUP motifs were similar to rRNA structure
432 motifs. To experimentally validate SINEUP–rRNA interactions we used PARIS2 and observed a few
433 instances of chimeric interactions, although we saw none to the helix 34 of 18S rRNA. The sparsity of

434 such SINEUP–rRNA interactions may be due to a limitation of psoralen cross-linking, which has bias for
435 staggered uridines and may not access RNA structures positioned inside complex ribonucleoprotein
436 regions[54]. Alternatively, the efficiency of PARIS2 to capture RNA–RNA interactions may be limited by the
437 level of expression of RNAs as well as by the brief half-lives of chimeric interactions[40,54]. Nevertheless,
438 our PARIS2 data identifies for the first time a direct interaction of SINEUP-GFP (SINEB2) RNA and
439 miniSINEUP-GFP (FRAM) RNA with both 18S and 28S rRNA, which is consistent with previous polysome
440 profiling data[38]. Even though we capture long chimeric interactions, in terms of sequence
441 complementarity we unambiguously identify only short discontinuous stretches of base-pairing between
442 SINE-derived RNA and rRNA. However, this may still be sufficient to influence translation, given that
443 circular IRESs have 7-mer–long complementarity to 18S rRNA[42] and HCV IRES needs only 3 nucleotide-
444 long Watson-Crick base-pairing with 18S rRNA to induce translation[50]. Moreover, RNA–RNA interactions
445 in 3D depend heavily on non-Watson-Crick base pairing and non-covalent interactions, such as base-
446 stacking and tertiary contacts by purine to minor-groove and tetraloop to tetraloop receptor[60], which
447 are not easily predicted. Therefore, we decided to take a closer look at the SINE-rRNA chimeras.

448 Notably, we discovered an interaction between a region comprising helix 14 and 15 of 18S rRNA (Int 1)
449 and motif 1 and 2 of SINEB2 RNA. FRAM RNA motifs 1 and 2 along with motifs 5 and 6 also interacted
450 with helix 15 of 18S rRNA. Helix 14 of 18S rRNA is known to interact with ribosomal protein rpL23[61]. In
451 addition, a protein ABCE1 (ATP Binding Cassette Subfamily E Member 1), which is involved in translation
452 initiation and termination, interacts with helix 14 during translation initiation[62]. Moreover, helices 14
453 and 15 bind with eukaryotic initiation factor 5b (eIF5b), a ribosome-dependent GTPase that mediates
454 joining of 40S and 60S subunits of ribosomes[63]. This indicates the possibility that SINEUP motifs 1 and 2
455 are involved in the joining of ribosomal subunits during translation initiation. Furthermore, we observed
456 contact of the region between motif 5 and 6 of SINEB2 RNA with domain III helices 55, 56, and 59 of 28S
457 rRNA. Helix 59 is a part of the ribosomal tunnel exit site, from where the nascent peptide emerges out of
458 the ribosome[64]. It suggests that SINEB2 RNA not only takes part in initiating target mRNA translation but
459 also in elongation. In addition, SINEB2 RNA motif 5 and the surrounding region was found to bind with
460 domain I, helix 25, expansion segments ES7c and ES7d of 28S rRNA. The connection with ES7c was also
461 found to be conserved in FRAM RNA, in which the SL1 region (containing a partial match to motif 5)
462 participated in the interaction. ES7 is the largest expansion segment positioned at the back of the 60S
463 ribosomal subunit. This highlights the importance of SINEUP structure motifs in their interaction with
464 the large subunit of the ribosome. We also observed chimeric interaction of FRAM RNA SL1 with 18S
465 rRNA helices 21, 37, and 38. Because the helix 38–42 region is crucial for mRNA binding and placement

466 of anticodon arms for tRNAs[65], this interaction is notable for the proper positioning of SINEUP target
467 mRNA and translation initiation. FRAM RNA also interacted with domains I, III, IV, and V of 28S rRNA.

468 These contacts at multiple rRNA sites indicate the dynamic changes in SINE RNA–rRNA interaction that
469 occur as translation progresses and highlight the role of SINEUPs in regulating both translation initiation
470 and elongation. It is to be noted that although we captured the dynamics of SINE RNA structure and
471 interaction with rRNA and RBP, many of these processes are very transient in nature and it is unlikely to
472 obtain a complete view of all intermediate RNA structures and interaction kinetics in possibly very short
473 time windows by existing in vivo methods. Considering this, further 3D analysis of the tertiary
474 connections of SINEUP RNA with rRNA and translation complexes by using cryogenic electron
475 microscopy or other approaches is needed in the future to construct a full 3D model of larger molecular
476 complexes. The 3D models of 18 SINE-derived RNAs in this study identified stabilizing stacking
477 interactions within structure motifs and demonstrated strikingly similar dynamics of mouse and human
478 SINE-derived RNAs in 3D space in the subcellular compartments. We list all the base-stacking and non-
479 canonical interactions determined from these models, which will provide a foundation for future
480 experimental 3D analysis.

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

492

493 **Data availability**

494 The icSHAPE sequencing data were deposited in Gene Expression Omnibus (GEO) with accession
495 numbers GSE146407 (for whole cell icSHAPE libraries of SINEUP-GFP) and GSE224534 (for whole cell

496 icSHAPE libraries of miniSINEUP-GFP and nuclear-cytoplasmic fractionated icSHAPE libraries). The
497 PARIS2 sequencing data and SINEUP RNA modification data are accessible through GEO accession
498 numbers GSE224533 and GSE224018 respectively. Nuclear-cytoplasmic fractionated seCLIP sequencing
499 data is available through the GEO accession number GSE227250.

500

501 **Acknowledgements-**

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

509

510 **Funding**

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

515

516 *Conflict of interest statement.* S.G. and P.C. are inventors on patent US9353370B2 and related
517 applications in Europe, USA and Japan, and H.S., H.T., S.G., and P.C. are inventors on patent application
518 102018000002411 and the related application PCT/IB2019/050914 held by SISSA in Italy and TranSINE
519 Therapeutics in the UK, of which S.G. and P.C. are founders and H.T. owns shares.

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

524

525 **Materials and Methods**

526 **Plasmids and cloning**

527 SINEUP-GFP (pcDNA3.1-based) synthetic SINEUP containing antisense *Uchl1* SINEB2[10] was used to
528 construct long SINEUP effector domain–swapping mutants. To swap the antisense *Uchl1* inverted
529 SINEB2 of SINEUP-GFP with other SINEB2 sequences of interest, two restriction sites for SacII and ClaI
530 restriction enzymes were introduced at the start and end of the SINEB2. For this purpose, a QuikChange
531 II Site-Directed Mutagenesis Kit (Agilent, catalog #200523) was used. All the mutagenesis primers were
532 designed using QuikChange Primer Design Program (Agilent) and are listed in **Supplementary Table 2a**.
533 The SINEB2 regions were PCR-amplified from RIKEN mouse cDNA clones AK143784 (antisense to *Txnip*)
534 and AK054076 (antisense to *Gadd45α*), then digested by the SacII and ClaI restriction enzymes and sub-
535 cloned in SINEUP-GFP to create swapping mutants of antisense *Txnip* and antisense *Gadd45α* SINEB2-a
536 and -b respectively. Antisense *Gadd45α* SINEB2-ab and ba, and all the full constructs of antisense *Uxt*
537 SINEB2–swapping mutants were obtained from Genewiz. The highly active antisense *Uchl1*_Δ5′-32nt
538 mutant[30] was included in screening as a positive control. Direct and deleted antisense (ΔAS) *Uchl1*
539 SINEB2 mutants were created by a conventional restriction digestion, PCR, and cloning method, and the
540 SINEUP mutant with deleted binding domain (ΔBD) was purchased from TransSINE Technologies.
541
542 Previously described miniSINEUP-GFP[29] plasmid consisting of antisense *Uchl1* SINEB2 as the effector
543 domain and −40/+4 binding domain design with respect to the start codon of GFP mRNA was used as a
544 backbone for miniSINEUP effector-domain swapping. A conventional PCR amplification, restriction
545 digestion, ligation, and cloning strategy was used to generate miniSINEUP-GFP mutants shown in **Fig. 1e**.
546 The SINEB2 region of RIKEN mouse cDNA clones AK029359 (from 774–960 bp for antisense *Uxt-b*
547 SINEB2), AK032194 (from 769–982 bp for antisense *Nars2* SINEB2), AK041742 (from 2334–2533 bp for
548 antisense *Abhd11* SINEB2), and AK048309 (from 1203–1416 bp for antisense Epb4.9 SINEB2) was
549 amplified by PCR (all primer sequences are shown in **Supplementary Table 2b**). To prepare the effector
550 domain for antisense *Uchl1* SINEB2 + antisense *Uxt-b* SINEB2 (**Fig. 1e**), the SINEB2 regions of RIKEN
551 cDNA clones AK078321 (from 521–690 bp for antisense *Uchl1* SINEB2) and AK029359 (from 774–960 bp
552 for antisense *Uxt-b* SINEB2) were first amplified by PCR separately and then ligated to make one
553 effector-domain insert (all primer sequences are shown in **Supplementary Table 2b**). Similarly, in the
554 case of 3× antisense *Uchl1* SINEB2, PCR-amplified SINEB2 from AK078321 (from 521–690 bp) was
555 consecutively ligated twice to generate three repeats of the same SINEB2. The rest of the SINEB2
556 effector domains originating from various mouse natural antisense lncRNAs (**Fig. 1e**, and **Supplementary**
557 **Table 2**) and the consensus sequence for SINEB2 subfamily B3 from the Repbase database[27,46] were
558 commercially obtained from Hokkaido System Science Co., Ltd. The backbone plasmid, miniSINEUP-GFP,
559 and amplified/commercial SINEB2 effector-domain inserts were digested by EcoRI and HindIII restriction
560 enzymes, then ligated and cloned.
561 The commercially available pEGFP-C2 (Clontech) plasmid was used as a sense EGFP construct. The
562 miniSINEUP-GFP FRAM plasmid design has been previously described[31].
563

564 **Cell culture and transfection**

565 Human embryonic kidney (HEK293T/17) cells (CRL-11268, ATCC) were grown in a 10-cm culture dish in
566 DMEM (1×) + GlutaMAX-1 (Thermo Fisher Scientific, 10569-044) supplemented with 10% fetal bovine
567 serum (Sigma, 172012) and 1% penicillin streptomycin solution (Wako, 168-23191) at 37°C and 5% CO_2
568 for 3–5 days. 70–90% confluent cells were treated with 0.05% w/v Trypsin-0.53 mmol/L EDTA 4Na
569 Solution with phenol red (Wako, 202-16931) for 5 min at 37°C to detach the cells, and 0.5 × 10^6 cells

570 were seeded in each well of 6-well plates (Falcon). After 24 h, cells were transfected with pEGFP-C2 and
571 SINEUP-GFP (long) or miniSINEUP-GFP swapping mutant plasmids at a molar ratio of 1:4.3 for long
572 SINEUPs GFP [0.19 pmol pEGFP-C2 + 0.81 pmol of long SINEUP] and 1:4.6 for miniSINEUP-GFP [0.19
573 pmol pEGFP-C2 + 0.87 pmol of miniSINEUP-GFP] using 10 µL of Lipofectamine2000 (Thermo Fisher
574 Scientific, 11668-019). After 24 h of transfection, cells were harvested.

575

576 **Protein extraction and Western blot**

577 Cells were washed with D-PBS (-) (Wako, 045-29795) and lysed using 140 µL of 10× Cell Lysis buffer (Cell
578 Signaling Technology, 9803S) with PMSF (Cell Signaling Technology, 8553S) per well in a 6-well plate.
579 Samples were rotated at low speed for 1 h at 4°C and then centrifuged (20,000 × *g*) for 10 min at 4°C.
580 The supernatant was collected, and the protein concentrations were measured by DC Protein Assay
581 (BioRad, 500-0116). The absorbance (750 nm) was measured by Multimode Plate Reader ARVO X3
582 (PerkinElmer). For Western blotting, 20 µg of extracted proteins mixed with 2× SDS sample buffer were
583 separated on 10% SDS PAGE gel (Mini PROTEAN TGX Precast Gel, 10%, 12-well comb; BioRad, 456-1035)
584 and transferred to a nitrocellulose membrane (Amersham Protran Premium ECL 0.45 µm; GE Healthcare
585 Life Sciences, 10600013) through semi-dry transfer at 25 V for 30 min using Trans-BLOT SD Semi-dry
586 Transfer Cell (BioRad). Tris-glycine buffer containing 20% methanol was used for transfer. The
587 membranes were blocked with 5% nonfat dry milk (Cell Signaling Technology, 9999S) in Tris Buffered
588 Saline with Tween-20 (10× TBST, Cell Signaling Technology, 9997) for 30 min with shaking. Proteins were
589 immunoblotted with primary antibodies followed by horseradish peroxidase– (HRP) conjugated
590 secondary antibodies. EGFP was detected by anti-GFP rabbit serum A-6455 (Life Technologies), ACTINB
591 was detected by monoclonal anti-β-actin antibody (Sigma Aldrich, A5441). Protein signals were detected
592 by ECL Western Blotting Detection Reagent (GE Healthcare Life Sciences, RPN2109) with FUJI LAS-3000
593 system (Fujifilm) and FUSION (Vilber-Lourmat). The band intensities were analyzed by Image J version
594 1.48 software (National Institutes of Health).

595

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

597 HEK293T/17 cells were washed with D-PBS (-) (Wako, 045-29795) and detached using 0.05% w/v trypsin
598 (Wako, 202-16931). Cells were collected by centrifugation for 5 min (6000 × *g*) at 4°C, and total RNAs
599 were subsequently extracted using an RNeasy Mini Kit (Qiagen, 74104). Extracted RNA samples were
600 processed through three rounds of DNase I digestion using a TURBO DNA-free Kit (Invitrogen, AM1907).
601 RNA quality was checked on Agilent 2100 Bioanalyzer (Agilent Technologies) using an Agilent RNA 6000
602 Nano Kit. Then, 500 ng of the total RNA was used for cDNA synthesis using the Prime Script 1st Strand
603 cDNA Synthesis Kit (TaKaRa, 6110A). Random 6-mer primers (0.3 µL) and 1.7 µL of oligodT primers were
604 mixed to capture RNAs. 2× diluted cDNA (1 µL) was used as a template for the quantitative real-time
605 PCR (qRT-PCR). For qRT-PCR, a SYBR Premix Ex Taq (Tli RNaseH Plus) Kit (TaKaRa, RR420S) was used.
606 Each reaction contained 1 µL of cDNA, 0.4 µL of reverse primer (10 µM), 0.4 µL of forward primer (10
607 µM), 10 µL of SYBR premix ExTaq, 0.4 µL of ROX reference dye (50×), and 10 µL of nuclease-free water
608 (reaction volume was 20 µL). EGFP and SINEUP-GFP levels were normalized by GAPDH. qRT-PCR was
609 performed by StepOnePlus and 7900HT Real-Time PCR Systems (Applied Biosystems) under the
610 following conditions: hold stage (1 cycle) for 30 s at 95°C, cycling stage (40 cycles) for 5 s at 95°C and for

611 30 s at 60°C, melt curve stage. Reverse transcription samples were taken to check for any plasmid DNA
612 contamination in cDNA. Each sample was performed in technical triplicates (n = 3, C_T standard deviation
613 > 0.2) and with biological replicates (n ≥ 3). 2^-ΔΔCt ±SD was analyzed by StepOne v2.3 and SDS v2.4
614 software (Thermo Fisher Scientific). The qRT-PCR primer sequences were as follows:

615 hGapdh_Fw: TCTCTGCTCCTCCTGTTC

616 hGapdh_Rv: GCCCAATACGACCAAATCC

617 EGFP_Fw: GCCCGACAACCACTACCTGAG

618 EGFP_Rv: CGGCGGTCACGAACTCCAG

619 SINEUP-GFP_Fw: CTGGTGTGTATTATCTCTTATG

620 SINEUP-GFP_Rv: CTCCCGAGTCTCTGTAGC

621 miniSINEUP-GFP_Fw: TTAATACGACTCACTATAGGGAGAC

622 miniSINEUP-GFP_Rev: CCGCATCTAGTGAACCGTCA

623

624 **Phylogenetic analysis and hierarchical clustering**

625 MEGA version 5 software[66] was used for sequence-based alignment, and mouse SINEB2 sequences
626 tested as effector domains in long and mini SINEUPs were aligned by ClustalW. SINEB2 sequence–based
627 molecular phylogenetic analysis was done using the Maximum Likelihood method. The evolutionary
628 history was inferred by the Maximum Likelihood method based on the Tamura 3-parameter model.
629 Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ
630 algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood
631 approach and then selecting the topology with superior log likelihood value. A discrete Gamma
632 distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter =
633 4.1774)). The tree was drawn to scale, with branch lengths measured in the number of substitutions per
634 site.

635

636 **icSHAPE library preparation**

637 The icSHAPE standard protocol was followed[12,48]. Briefly, HEK293T/17 cells were co-transfected with
638 pEGFP-C2 and SINEUP-GFP or miniSINEUP-GFP swapping mutants using Lipofectamine2000 reagent in 6-
639 well plates. Cells were harvested after 24 h of transfection, pooled together, and treated with NAI-N_3
640 reagent (modified) (NAI-N3 reagent was received as a gift from Dr. Howard Chang's lab at Stanford
641 University) for in vivo SHAPE modification or with DMSO (mock control). Two biological replicates for
642 each of the NAI-N_3 and DMSO libraries were made. RNA was extracted using an RNeasy Mini Kit (Qiagen,
643 74104) and digested with DNaseI to remove DNA contamination using a TURBO DNA-free Kit (Invitrogen,
644 AM1907). RNA was ribo-depleted using a magnetic RiboZero kit (Illumina, MRZH11124), and RNA quality
645 was checked by Agilent 2100 Bioanalyzer (Agilent Technologies) nano and pico chip. Next, 500 ng of
646 ribo-depleted RNA was biotin-clicked to label-probed bases followed by RNA fragmentation for 40 s
647 using RNA fragmentation reagents (Thermo Fisher Scientific, AM8740). Fragmented RNA was end-

648 repaired, and the icSHAPE adapter ligation standard protocol was followed[48] (adapter sequences in
649 **Supplementary Table 5a**). Next, size selection was performed, and RNA was reverse-transcribed using
650 icSHAPE-specific barcoded reverse-transcription primer followed by streptavidin-based selection and
651 size selection of truncated cDNA products. Purified cDNA was circularized using CircLigase II enzyme
652 (Epicentre, CL9025K), then library PCR, size selection, and quantification were performed. Resultant
653 libraries were sequenced on a HiSeq2500 platform (Illumina, 100 bases, single-end). All the oligos used
654 in the icSHAPE library preparation are noted in **Supplementary Table 5a**.

655

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

657 The standard icSHAPE bioinformatics pipeline was used[48]. Briefly, after removing PCR duplicates and
658 adapter trimming, raw sequencing data were mapped to EGFP and SINEUP transcripts index using
659 Bowtie2. Next, transcript abundance and reverse transcription (RT) stop numbers were calculated
660 before calculating the correlation between replicates and replicate merging. The mean of the RT stops of
661 the 90–95% most reactive bases was normalized as 1 and all other RT stops were scaled proportionally.
662 Then enrichment reactivity scores were calculated by subtracting the background signal (DMSO data),
663 and outliers were removed by 90% Winsorization (in which the top 5th percentile is set to 1 and the
664 bottom 5th percentile is set to 0). Valid enrichment reactivity scores were filtered and visualized in
665 integrated genome viewer (IGV v2.11.1). The icSHAPE enrichment scores were used as a soft constraint
666 in RNAfold (ViennaRNA Package)[67]. Linear mapping method was used to derive pairing probabilities, and
667 a method described by Zarringhalam *et al.* (2012)[68] was used to incorporate guiding pseudo energies
668 into the folding algorithm. To draw the secondary structure, RNAcanvas[69] (https://rna2drawer.app/) was
669 used.

670 Sequence and icSHAPE-guided 2D structure of all the SINEB2s were pairwise compared with the
671 structure and sequence of antisense *Uchl1* SINEB2 using ExpaRNA tool[49], which computes the exact
672 matching sequence-structure motifs common to two RNAs. Thus, derived motif files from all the
673 pairwise comparisons were then screened to extract the positions of Uchl1 SINEB2s that were present in
674 one or multiple motifs in all of the comparisons. Since SINEB2 sequences are highly variable, in most
675 cases only a partial sequence match could be found. Due to their repeating sequence composition, in
676 some instances one query sequence was matched to multiple motifs, or one motif matched to multiple
677 positions in the query. In that case, the longer matching stretch was preferred, and different motif
678 numbers were assigned based on their sequence and structure neighborhood and position on SINEB2.
679 Additionally, SINEUP 2D motifs were compared with existing 3D motifs in the RNA Characterization of
680 Secondary Structure Motifs database (CossMos)[51].

681

682 **Nuclear-cytoplasmic fractionation for icSHAPE**

683 HEK293T/17 cells were co-transfected with 0.8 pmol SINEUP-GFP or miniSINEUP-GFP FRAM (0.8 pmol)
684 and 0.2 pmol sense pEGFP-C2 plasmids (Clontech) in each well of a 6-well culture plate. Four 6-well
685 plates for each sample were prepared. At 24 h post-transfection, the medium was aspirated, and the
686 cells were washed and collected in 1 mL of 1× D-PBS (-) (Nacalai tesque, 14249-24) per well. Cells from
687 one 6-well plate (~1×10^7 cells) were pooled together in a 15-mL tube and centrifuged at 300 × *g* for 2

688 min at room temperature. The resultant supernatant was removed, and the pelleted cells were
689 resuspended in either 250 µL of 100 mM NAI-N3 reagent (Enamine, EN300-247313) for modification or
690 in 5% DMSO (Sigma Aldrich, D2650) solution in 1× PBS (Gibco) for mock treatment. The reaction was
691 incubated on a rotating wheel at 37°C for 5 min at room temperature and then stopped by
692 centrifugation at 2500 × *g* for 1 min at 4°C and further removal of the supernatant. Next, the protocol by
693 Pandya-Jones and Black (2009)[70] was utilized for cellular fractionation. In brief, the NAI-N3- and DMSO-
694 treated cell pellets were resuspended in ice-cold NP-40 lysis buffer (10 mM Tris-HCl pH 7.5, 0.15% NP-
695 40, 150 mM NaCl), then the lysate was added on top of 2.5 volumes of chilled 25% sucrose cushion in
696 lysis buffer without NP-40 followed by centrifugation at 19,000 × *g* for 10 min at 4°C. The resultant
697 supernatant containing the cytoplasmic fraction was collected. The pellet was further gently rinsed with
698 200 µL chilled 1× PBS/1 mM EDTA and centrifuged at 9,730 × *g* for 2 min at 4°C to remove the
699 supernatant. Next, to extract the nuclear fraction, the protocol by Brugiolo *et al.* (2017)[71] was followed.
700 The pellet was resuspended in 100 µL nuclear buffer I (50% glycerol v/v, 20 mM Tris-HCl pH 7.9, 75 mM
701 NaCl, 0.5 mM NaCl, 0.85 mM DTT, 0.125 mM PMSF) and then in 900 µL nuclear buffer IIA (20 mM HEPES
702 pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 7.5 mM $MgCl_2$, 1M Urea, 1% NP-40, 400 U of RNase
703 inhibitor). The solution was mixed thoroughly by vortexing and then pipetting up and down, and then
704 incubated on ice for 20 min. Finally, the nucleoplasmic fraction was extracted by centrifuging at 14,000 ×
705 *g* for 5 min at 4°C and collecting the supernatant. The success of fractionation was checked by Western
706 blot and qRT-PCR for known nuclear and cytoplasmic marker genes. Four nucleoplasmic or cytoplasmic
707 fractions, each coming from four 6-well plates of the same sample, were pooled together for further
708 RNA extraction and icSHAPE library preparation as described above. All the steps are the same as in the
709 original protocol[48], except the removal of free 3' RNA linkers after ligation, for which the completed
710 ligation reaction was treated with 4 µL $RecJ_f$ (New England Biolabs, M0264L) and 3 µL 5′ Deadenylase
711 (New England Biolabs, M0331S) at 37°C for 1 h and further cleaned up by using a Zymo RNA Clean and
712 Concentrator-5 purification kit (Zymo Research, R1016). All the oligos used in the library preparation are
713 noted in **Supplementary Table 5a**.

714

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

716 Approximately 0.5 × 10^7 HEK293T/17 cells were plated on a 10-cm culture dish and after 24 h,
717 transfected with SINEUP-GFP and pEGFP-C2 plasmids as described above. After 24 h of transfection, 2 ×
718 10^7 cells were gently washed with 1× PBS and then crosslinked by 400 mJ/cm, 254 nm UV light
719 irradiation on ice with the lid off for 15 min. Next, the cells were collected by scraping in 1 mL chilled
720 PBS followed by centrifugation at 200 × *g* for 5 min at 4°C. The resultant supernatant was discarded, and
721 the pellet was used for nuclear-cytoplasmic fractionation following the protocol by Brugiolo *et al.*
722 (2017)[71]. The pellet was first resuspended in 2 mL hypotonic buffer (10 mM Tris-HCl 7.5 pH, 10 mM KCl,
723 1.5 mM $MgCl_2$, 0.5 mM DTT, 1:200 proteinase inhibitor just prior to use) and incubated on ice for 5 min
724 followed by centrifugation at 500 × *g* for 10 min at 4°C. The supernatant containing hypotonic fraction
725 was discarded, and the pellet was resuspended in 1 mL lysis buffer 0.3 (50 mM Tris-HCl 7.5 pH, 150 mM
726 NaCl, 2 mM $MgCl_2$, 0.3% NP-40, 1:200 proteinase inhibitor just prior to use) and incubated on ice for 10
727 min. The mix was then centrifuged at 1000 × *g* for 5 min at 4°C, and the supernatant containing the
728 cytoplasmic fraction was collected in a new 1.5-mL tube. The pellet was mixed with 1 mL lysis buffer 0.5
729 (50 mM Tris-HCl 7.5 pH, 150 mM NaCl, 2 mM $MgCl_2$, 0.5% NP-40, 1:200 proteinase inhibitor just prior to
730 use) and incubated on ice for 10 min followed by centrifugation at 1000 × *g* for 5 min at 4°C. The

731 supernatant was discarded, and the pellet was first resuspended in 100 µL nuclear buffer I (as used for
732 icSHAPE fractionation, without addition of PMSF) and then in 900 µL nuclear buffer IIA (as used for
733 icSHAPE fractionation) with 1:200 proteinase inhibitor just prior to use. The pellet was mixed well by
734 vortexing and then pipetting up and down, and then incubated on ice for 15 min. The reaction mix was
735 then centrifuged at 15,000 × *g* for 5 min at 4°C, and the nucleoplasm containing supernatant was
736 transferred to a new 1.5-mL tube. All fractions were separately centrifuged at 20,000 × *g* for 5 min at 4°C
737 and tested for the presence of markers specific to nuclear or cytoplasmic fractions by Western blot.

738 Next, the seCLIP protocol described by Van Nostrand *et al.* (2017)[45] was utilized to prepare nuclear and
739 cytoplasmic seCLIP libraries. In brief, nucleoplasmic and cytoplasmic lysates were treated with 4 U of
740 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.
741 The reaction was then cooled on ice, 7 µL Murine RNase Inhibitor (New England Biolabs, M0314) was
742 added, and the mix was centrifuged at 15,000 × *g* for 15 min at 4°C. The resultant supernatant was
743 mixed with 15 µg of anti-hnRNPK antibody (Abcam, ab39975) coupled to 200 µL magnetic beads
744 (Dynabeads M-280 Sheep Anti-mouse IgG, Invitrogen, 11201D) and incubated at 4°C overnight. Around
745 2% of the input samples (containing all RNA-protein complexes) were saved before
746 immunoprecipitation. Immunoprecipitated samples were first end-treated with FastAP thermo-sensitive
747 alkaline phosphatase (Thermo Scientific, EF0651) and T4 poly nucleotide kinase (New England Biolabs,
748 M0201L) and then ligated with 3′ RNA linker (with sample barcodes, see sequence in **Supplementary**
749 **Table 5b**). Next, both immunoprecipitated and input RBP–RNA complexes were visualized by Western
750 blot imaging and the region corresponding to HNRNPK and above (from 55–180 kDa) was sliced from the
751 preparative nitrocellulose membrane blot. The RNA was extracted by treating the membrane slices with
752 urea and proteinase K followed by acid phenol–chloroform extraction and purification by using a Zymo
753 RNA Kit (R1013). Further treatment of samples and cDNA library preparation are the same as described
754 in the original protocol[45]. Libraries were sequenced on the HiSeq2500 platform (150 cycles, single-end).

755

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

757 Fastq files were processed with two rounds of cutadapt (which trims off adapters and controls for
758 double ligation events) before being mapped to a custom genome (hg38 + SINEUP construct) using STAR
759 v2.7.8a with parameters as in the ENCODE eCLIP project pipeline
760 (https://www.encodeproject.org/eclip/) (namely, --outFilterMultimapNmax 1, --
761 outFilterMultimapScoreRange 1, --outFilterScoreMin 10, --alignEndsType EndToEnd). Reads mapped on
762 the sense strand of the SINEUP construct were removed. Peaks were called on mapped reads using
763 clipper v0.2.0 (https://github.com/YeoLab/clipper) with custom data files (gff, exons bed file) containing
764 the SINEUP region, with default parameters. Scripts from the merge_peaks github repository
765 (https://github.com/YeoLab/merge_peaks) were used to normalize seCLIP reads over input reads for
766 each replicate (overlap_peakfi_with_bam.pl) before merging input-normalized peaks into a final set of
767 significant binding regions (compress_l2foldenrpeakfi_for_replicate_overlapping_bedformat.pl).

768

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

770 Transcribed SINEUP-GFP RNA was synthesized in vitro as described by Toki *et al.* (2020)[72] by using a
771 mMESSAGE mMACHINE SP6 Transcription Kit (Thermo Fisher Scientific) and was then modified by
772 following the protocol by Mandal and Rossi (2013)[73]. For the 20% Ψ modification, UTP was replaced with
773 20% pseudouridine-5′-triphosphate (TriLink, USA; final concentration, 7.5 mM) by mixing the reagents
774 with 40 ng/µL of linearized SINEUP RNA. To add a poly-A tail, 1–10 µg of RNA transcribed in vitro was
775 treated with Escherichia coli poly(A) polymerase (5000 U/mL, New England Biolabs, M0276) at 37°C for
776 30 min followed by clean-up using an RNeasy Mini kit (Qiagen) and the modified transcripts of SINEUP-
777 GFP RNA were recovered.

778

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

780 HEK293T/17 cells were plated on a 10-cm dish and SINEUP-GFP plasmids were transfected for 24 h; then
781 2×10^7 cells were harvested by Trizol/chloroform extraction, and total RNAs containing SINEUP-GFP
782 RNAs at the aqueous phase were extracted by using an RNeasy mini Kit (Qiagen). The RNA was then
783 digested with 4 U of TURBO DNase (TURBO DNA-free kit, Invitrogen, AM1907) at 37°C for 90 min
784 followed by DNase inactivation.

785

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

787 The total RNA volume was adjusted to 100 µL with lysis buffer (50 mM Tris-HCl pH 7.0, 10 mM EDTA, 1%
788 SDS, 1:100 RNase inhibitor) and mixed with two volumes of hybridization buffer (750 mM NaCl, 50 mM
789 Tris-HCl pH 7.0, 1 mM EDTA, 1% SDS, 15% formamide, 1:200 RNase inhibitor) with additional 6.6%
790 formamide. Then, 100 pmol of SINEUP-GFP-specific total probe (2 µL of 50 µM probe) was added to this
791 mix (probe design as described by Toki *et al.*, 2020)[38], and the reaction was incubated at 37°C overnight
792 with vertical rotation. Next, 100 µL of MagCapture Tamavidin 2-REV magnetic beads (WAKO) were
793 washed and resuspended in 100 µL lysis buffer. This bead suspension was added to the completed
794 hybridization reaction and incubated at 37°C for 30 min with mixing. The solution was then magnetically
795 separated, and the supernatant was discarded. The beads were washed four times with pre-warmed
796 wash buffer (2× saline sodium citrate, 0.5% SDS, 1:200 RNase inhibitor), and the supernatant was
797 discarded. After this, the beads were resuspended in 100 µL ProK buffer (10 mM Tris-HCl pH 7.5, 0.8%
798 SDS, 15 mM EDTA) without proteinase K and incubated at 65°C for 5 min with end-to-end mixing at 450
799 rpm. The first pull-down RNA was then extracted by the Trizol/chloroform method and purified by
800 miRNeasy mini column (Qiagen, 217004). This pull-down and RNA extraction was repeated once again to
801 enrich SINEUP-GFP RNAs.

802

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

804 Libraries of the direct SINEUP-GFP RNA transcribed in vitro and in-cell were prepared with an SQK-
805 RNA002 kit (Oxford Nanopore Technology). The original reverse-transcription adapter (RTA) was used
806 for in vitro samples and four barcoded RTA adapters described by Leger *et al.* (2021)[74] were used for in-
807 cell samples. The libraries were applied to MinION Mk1C sequencer (Oxford Nanopore Technologies)
808 and sequenced for 72 h.

809

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

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

827

828 **PARIS library preparation**

829 HEK293T/17 cells were transfected with pEGFP-C2 and SINEUP-GFP (SINEB2) or miniSINEUP-GFP FRAM
830 plasmids in 6-well plates as described above. After 24 h of transfection, cells were gently washed with
831 1× PBS and treated with 150 µL 1× PBS (control) or 1 mg/mL Amotosalen (MedChemExpress, HY-
832 107004A) crosslinking solution added to each well of the 6-well plates, followed by incubation under
833 standard cell-culture conditions (37°C, 5% CO_2) for 30 min. Next, cells were irradiated for 30 min by 365-
834 nm UV light while on ice with swirling of the plate every 10 min. Then, the crosslinking solution was
835 removed, and the cells were scraped off into 1 mL chilled 1× PBS in each well and transferred to a 15-mL
836 tube. Crosslinked cells were pelleted by centrifuging at 300 × *g* for 5 min at 4°C, and the supernatant
837 was discarded. Next, for the total nucleic acid (TNA) extraction and PARIS library preparation, the
838 protocol by Zhang *et al.* (2021)[40] was used with some modifications. Briefly, the cell pellet was lysed by 6
839 M GuSCN with pipetting followed by treatment with 0.5 M EDTA and passed through a 25G needle 20
840 times. After this, the proteins were digested by proteinase K (final concentration 1 mg/mL) at 37°C for 1
841 h on a thermomixer with shaking, and solution was again passed through the 25G needle. After TNA was
842 precipitated by phenol-isopropanol precipitation as described in Zhang *et al.* (2021)[40], 100 µg of TNA
843 sample was treated with 50 units of TURBO DNase (Invitrogen, AM2239) at 37°C for 20 min. RNA was
844 precipitated using phenol-isopropanol and further purified using RNA clean and concentrator kit (Zymo
845 Research, R1015). The RNA quality was measured with Bioanalyzer Nano Kit (Agilent). DNase-treated
846 RNA (10 µg) was fragmented by using 5 units of ShortCut RNase III (New England Biolabs, M0245) for 5
847 min at 37°C and the reaction was stopped by adding 10 µL nuclease free water, 100 µL of RNA binding
848 buffer (Zymo Research, R1015) and 150 µL of 100% ethanol and then purified through a column from
849 the purification kit (Zymo Research, R1015) followed by quality check with a Bioanalyzer small RNA kit
850 (Agilent).

851 For 2D gel purification of RNA duplexes, 8% 1-mm thick denatured first-dimension gel was prepared
852 using the UreaGel System (National Diagnostics, EC-833). A 10-µL sample of fragmented RNA was run on
853 the first-dimension gel in 0.5x Novex TBE running buffer (Invitrogen, LC6675) at 30 W for 9 min. The gel
854 was visualized under 300 nm transillumination after staining with SYBR Gold (Invitrogen, S11494) and
855 each lane was sliced between 50 bp to topside from the first-dimension gel. Next, 16% 1-mm thick
856 denatured second-dimension gel was prepared using the UreaGel System (National Diagnostics, EC-833).
857 The second-dimension gel was poured to encapsulate the gel slices from the first-dimension gel, and
858 0.5x TBE buffer warmed to 60°C was used to facilitate the denaturing of crosslinked RNA. The gel was
859 run at 30 W for 50 min and then stained with SYBR Gold, and crosslinked RNA was sliced out from the
860 upper diagonal of the 2D gel. The gel slices were cut into smaller pieces and transferred to a 0.5-mL tube
861 with a hole drilled in the bottom with a 20G needle. This 0.5-mL tube was placed inside a 2-mL tube and
862 centrifuged at 16,000 x *g* for 5 min to crush the gel through the hole. Then 1.5 mL of crushing buffer (20
863 mM Tris-HCl pH 7.5, 0.25 M sodium acetate, 1 mM EDTA, 0.25% SDS) was added to the crushed gel and
864 rotated at room temperature overnight. After this, 500 µL of gel slurry was transferred to a Micro Spin
865 column (Cytiva, 27356501) and centrifuged at 3400 x *g* for 1 min or until all gel slurry was filtered. This
866 filtered gel was then passed through an Amicon Ultra 10K column (Millipore, UFC501096) and further
867 purified using the Zymo RNA Clean and Concentrator-5 column.

868 Next, in order to perform proximity ligation, 10 µL of the purified double-stranded RNA was treated with
869 7.5 unit/µL of T4 RNA Ligase I, high concentration (New England Biolabs, M0437) in a reaction mix
870 containing 1× T4 RNA Ligase Buffer (New England Biolabs, M0437M), 1U SUPERase•In RNase inhibitor
871 (Invitrogen, AM2694), and 0.5 mM ATP at room temperature overnight. The proximity-ligated RNA was
872 purified by using a Zymo RNA clean and concentrator-5 column.

873 To reverse crosslink the RNA, the samples were mixed with 3 µL of 25 mM acridine orange on ice and
874 irradiated with 254-nm UV light for 30 min. The reverse crosslinked sample was transferred to a new
875 1.5-mL tube and purified using a Zymo RNA Clean and Concentrator-5 column. Further, the reverse
876 crosslinked RNA was heated at 80°C for 90 s and then snap-cooled on ice. The adapter was ligated by
877 adding 14 µL of adapter ligation mix (1.5 units/µL T4 RNA ligase I high concentration, 1.5 µM adapter (5
878 Biotin), 10% DMSO, 12.5% PEG 8000, 5 mM DTT, 1× T4 RNA Ligase buffer) to 6 µL of reverse crosslinked
879 RNA and incubating the reaction for 3 h at room temperature. The sequences of all the oligos used in
880 the PARIS library preparation are listed in **Supplementary Table 5c**. After adapter ligation, free adapters
881 were removed by adding 3 units of $RecJ_f$ (New England Biolabs, M0264) and 1.7 units of 5 deadenylase
882 (NEB, M0331S) and incubating at 37°C for 1 h. The adapter-ligated RNA was purified using a Zymo RNA
883 Clean and Concentrator-5 column and eluted in 10 µL of nuclease free water. Next, 2 µL of custom
884 reverse transcription primer (with barcode, 1 µM, sequence in **Supplementary Table 5c**) and 1 µL of 10
885 mM dNTPs were added to the purified RNA and the sample was heated at 65°C for 5 min in a thermal
886 cycler followed by rapid cooling on ice. To this mix, 8 µL reverse transcriptase mix was added, which
887 contained 4 µL of 5x SSIV buffer (Mn2+) (250 mM Tris-HCl, 375 mM CH3COOK, 7.5mM $MnCl_2$), 10 mM
888 DTT, 1 unit/µL SUPERaseIn RNase inhibitor (Invitrogen, AM2694), 10 units/µL SuperScript IV (Invitrogen,
889 18090010). The reaction was incubated at 25°C for 15 min, 42°C for 10 h; hold at 10°C.

890 The reverse transcribed RNA was captured on streptavidin-coated magnetic beads. For this purpose, 10
891 µL of Dynabeads M-270 Streptavidin (Invitrogen, 65305) were washed in 20 µL LiCl buffer (7 M LiCl, 20
892 mM Tris-HCl (pH7.5), 2 mM EDTA, 0.1% Tween20) and then resuspended in 45 µL of LiCl buffer. This
893 beads suspension was added to the reverse transcribed RNA and incubated at 37°C for 30 min with

894 pipette mixing in between. The supernatant was removed by separation on a magnetic stand, and beads
895 were washed three times with 100 µL TE wash buffer (10 mM Tris-HCl (pH7.5), 1 mM EDTA, 0.1%
896 Tween20) and finally resuspended in 30 µL elution buffer (1x RNaseONE buffer, 1% Tween20, 10 µM P3-
897 tall primer). Next, the beads solution was heated at 95°C for 5 min then immediately cooled on ice and
898 magnetically separated to recover the released cDNA containing supernatant in a new 1.5-mL tube. The
899 beads were again resuspended in 30 µL elution buffer, separated on a magnet, and supernatant was
900 recovered and transferred to the released cDNA tube from the previous step. Next, the cDNA was
901 treated with 1 µL of RNaseONE (Promega, M4261), 1 µL of RNaseH (TaKaRa, 2150A), and 1 µL of RNase
902 cocktail (Invitrogen, AM2286) for 30 min at 37°C and further purified by using a Zymo DNA Clean and
903 Concentrator column (Zymo Research, D4033).

904 After this, 4 µL of circularization reaction mix containing 5 units/µL CircLigase II (Lucigen, CL9021K), 1×
905 CircLigase II buffer (Lucigen, CL9021K), and 2.5 mM $MnCl_2$ was added to 16 µL of cDNA and incubated at
906 60°C for 100 min followed by 80°C for 10 min. Circularized cDNA (1 µL) was used to determine the PCR
907 cycles required for the library amplification. For this, 39 µL of qPCR mix containing 1× Phusion mix (New
908 England Biolabs, M0531S), 0.125 µM P6 Tall v4 primer, 0.125 µM P3 Tall v4 primer, 0.15 µL ROX, 1/3800
909 SYBR (Invitrogen, S-7563) was added to 1 µL circularized cDNA and amplified at 95°C for 45 s; 20 cycles
910 of 95°C for 15 s, 65°C for 30 s, 72°C for 45 s on a StepOnePlus Real-Time PCR System (Applied
911 Biosystems, 4376598). The qPCR was stopped just before the amplification plot reached the plateau and
912 the corresponding cycle number was noted. For the final library amplification, 6 µL of circularized cDNA
913 was added to 44 µL PCR mix (1× Phusion mix, 0.15 µM P6 Tall v4 primer, 0.15 µM P3 Tall v4 primer) and
914 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
915 routine thermocycler. The PCR product was purified using Zymo DNA clean and concentrator column
916 (Zymo Research, D4013) and ran on 6% native TBE gel. DNA from 175 to 300 bp was excised, crushed,
917 and purified. Library quality was checked by using a Bioanalyzer High-sensitivity DNA chip and
918 concentration was quantified by using a qPCR KAPA library quantification kit (KAPABiosystems, KK4835).
919 Four barcoded libraries were pooled together and sequenced on HiSeq2500 platform (150 nt, single-
920 end, rapid mode).

921

922 **PARIS data analysis**

923 To check the SINEUP regions interacting with rRNAs, the PARIS2 analysis strategy was used as described
924 by Zhang *et al.* (2021)[40] (https://github.com/minjiezhang-usc/PARIS2). In short, the 3′ end adapters
925 were removed by Trimmomatic v0.39, and the sequencing reads were split according to the sample
926 barcodes using splitFastq script from icSHAPE pipeline[48]. Next, PCR duplicates were removed using the
927 readCollapse script from the icSHAPE pipeline[48] followed by removal of the 5′ header using
928 Trimmomatic v0.39. The reads were then mapped to a custom genome generated by adding SINEUP-
929 GFP (SINEB2), miniSINEUP-GFP (FRAM), and sense EGFP genes as separate chromosomes to the hg38
930 masked genome[40], which were previously engineered to contain a single copy of multicopy genes such
931 as rRNA and snRNAs. Mapping was done by STAR v2.7.10a following the parameters described in the
932 PARIS2 paper[40]. To extract SINEUP-rRNA chimeras, first the reads that were mapped to rRNA were
933 extracted by samtools (as in the PARIS2 analysis strategy) and then rRNA chimeric reads in which one of
934 the chimera arms was mapped to SINEUPs were extracted by using the grep command. The extracted
935 file was sorted by samtools and visualized on the integrated genome viewer (IGV v2.11.1). VfoldMCPX[75]

936 was used to predict the structure of the chimera. Human 18S rRNA structure (ID: URS00005A14E2) and
937 28S rRNA structure (ID: URS0000C873C2) were taken from the RNAcentral database
938 (https://rnacentral.org/).

939

940 **3D structure modelling**

941 SINE RNAs 2D structure information derived from icSHAPE was machine translated into 3D structure
942 models by RNAComposer[76] (https://rnacomposer.cs.put.poznan.pl/). The 3D models were visualized by
943 using Mol* tool[77] of RCSB PDB[78] (https://www.rcsb.org/) and analyzed by RNApdbee 2.0[79]
944 (http://rnapdbee.cs.put.poznan.pl/) to recognize the positions with base-stacking and non-canonical
945 interactions.

946

947 **Figure Legends**

948 **Figure 1. SINEB2 RNAs from different mouse antisense lncRNAs can act as SINEUP effector domain and**
949 **upregulate GFP protein expression in HEK293T/17.** (**a**) Schematic representation of the sense/antisense
950 EGFP system used to validate the effect of different SINEB2 RNAs on translation. The SINEUP binding
951 domain (BD) is shown in green. The inverted SINEB2 region is the effector domain (ED) shown in dark
952 gray. (**b**) SINEUP construct design to verify independent and combined effects of two SINEB2 elements
953 from the same antisense lncRNA. (**c**) Western blot–based EGFP band intensities normalized to ACTB
954 expression as a measurement of SINEUP-induced protein fold change observed in the HEK293T/17 cell
955 line at 24 h post-transfection. Empty vector was used as control. (**d**) Western blot images with anti-GFP
956 and anti-ACTB antibodies. ACTB was used as a loading control. Abbreviations: AS, antisense; Direct,
957 direct orientation of SINEB2; Δ, deletion; a and b, two different SINEB2 elements from the same
958 antisense lncRNA; ab and ba, the two SINEB2 elements in original and reversed order, respectively; AS
959 Uchl1 SINEB2–RS, antisense *Uchl1* SINEB2 construct without any restriction enzyme sites around the
960 SINEB2 elements. (**e**) GFP protein fold change in HEK293T cells after co-transfection with sense EGFP
961 and miniSINEUP (at 24 h post-transfection). Western blot images and corresponding GFP band
962 intensities (normalized to ACTB expression level) calculated using ImageJ software. An empty vector was
963 used as the negative control and antisense *Uchl1* SINEB2 containing miniSINEUP (shown as AS *Uchl1*) as
964 the positive control. Sample labels indicate names of antisense lncRNAs from which the SINEB2
965 sequences were isolated, which represent four sub-families (based on RepeatMasker annotation): B3,
966 B3A, Mm2, and Mm1t. SINEB2 elements are from the B3 sub-family unless specifically stated. AS *Uchl1* +
967 AS *Uxt*-b is the combination of AS *Uchl1* SINEB2 and the second SINEB2 of AS *Uxt*; 3× indicates three
968 repeats of SINEB2; and SINEB2/B3 consensus is the B3 sub-family consensus sequence taken from the
969 RepBase database. N = 3 (for c) and n = 5 (for e); error bars ± standard error mean; **** $P < 0.00005$;
970 *** $P < 0.0005$; ** $P < 0.005$; * $P < 0.05$ by two-tailed Student's *t*-test.

971

972 **Figure 2. In-cell SINE structures and conserved motifs in functionally active SINEUPs.** (**a**) Antisense (AS)
973 *Uchl1* SINEB2 RNA secondary structure based on icSHAPE data (right) is compared with the one
974 previously identified using chemical footprinting[11]. Squares highlight the common sequences and
975 structure motifs in in-cell structures of SINE, with their corresponding motif number. AGG (blue) and

976 UGG (red) motifs are shown. (**b**–**d**) Location of SINEUP motifs in icSHAPE-derived structures of
977 functionally active SINE RNAs from (**b**) antisense *Txnip* SINEB2, (**c**) antisense *Uxt* SINEB2-b and, (**d**)
978 human SINE element FRAM. (**e**) Effect of antisense *Gadd45α* SINEB2-a deletion mutants on GFP protein
979 expression as measured by Western blot. An empty vector was used as the negative control. Five
980 mutants with random deletion, as listed in the Figure, were compared with the full-length construct
981 (Full). N = 3, error bars ± standard deviation, * $P < 0.05$ by two-tailed Student's *t*-test. (**f**) Comparison of
982 global 2D structure and local SINEUP motifs (colors as in Figure 2a) of functionally dormant wild-type
983 antisense *Gadd45α* SINEB2-a (derived by icSHAPE) and its functionally active mutant antisense *Gadd45α*
984 SINEB2-a [Δ27–36] (predicted by RNA fold). AS, antisense; Δ, deletion; SL1, stem-loop 1. Nucleotide
985 color indicates normalized icSHAPE reactivity score: red, ≥ 0.8; yellow, from above 0.5 to 0.8; white
986 encircled with grey, ≤ 0.5; bases not encircled, no icSHAPE data available.

987

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

1000

1001 **Figure 4. Mouse SINEB2 RNA interacts with human rRNA in cell.** Antisense (AS) *Uchl1* SINEB2
1002 interacting regions (Int) in human (**a**) 18S rRNA (Int1) and (**b**) 28S rRNA (Int2 and 3) identified by PARIS2.
1003 Different interaction sites are color-coded, and rRNA-SINEB2 chimera are shown in the corresponding
1004 color. Possible base-pairing based on sequence complementarity of chimeric segments is shown. (**c**)
1005 SINEB2 regions interacting with 18S and 28S rRNA regions marked in (a) and (b) on icSHAPE data-driven
1006 SINEB2 structure from whole-cell lysate. SINEUP motifs from M1 to M7 are indicated by dashed squares.

1007

1008 **Figure 5. In-cell interaction of functional human SINE element FRAM with rRNA.** (**a**) Regions of 18S
1009 rRNA found to form chimera with the SINEUP FRAM region in PARIS2 data. (**b**) Corresponding binding
1010 regions to (a) are highlighted on FRAM RNA. (**c**) Description of FRAM-18S rRNA chimera and interaction
1011 sites marked in (a) and (b). (**d**) 28S rRNA regions identified as FRAM interaction sites by PARIS2 and (**e**)
1012 their corresponding binding regions on FRAM. (**f**) Details of FRAM-28S rRNA chimeric interactions
1013 highlighted in (d) and (e). Different interaction sites (Int) are represented by different colors. Possible
1014 base-pairing based on sequence complementarity of chimeric segments is shown in (c) and (f). In (b) and
1015 (e) the icSHAPE-guided FRAM structure from whole cell lysate is shown. Matching regions to SINEUP
1016 motifs from M1 to M7 are indicated by dashed squares.

1017

1018 **Figure 6. Machine-translated 3D RNA structure models based on in vivo SINE RNA 2D structures.** 3D
1019 structure models for (**a**–**d**) mouse antisense (AS) *Uchl1* SINEB2 RNA in the nucleus (**a**,**b**) and cytoplasm
1020 (**c**,**d**) and for (**e**–**h**) FRAM RNA in the nucleus (**e**,**f**) and cytoplasm (**g**,**h**). SINEUP structure motifs M1–M7
1021 are shown. Red frames show enlarged views of the SL1 region. Panels b, d, f, and h show the molecular
1022 surface view of the 3D models.

1023

1024 **Supplementary Table 1:** SINEB2 sequences tested as effector domain (ED) in (**a**) long SINEUP and (**b**) in
1025 MiniSINEUP-GFP.

1026

1027 **Supplementary Table 2:** (**a**) Primers used for QuikChange II site-directed mutagenesis. (**b**) Primers used
1028 for PCR-cloning of miniSINEUP-GFP constructs.

1029

1030 **Supplementary Table 3:** List of base-stacking positions in SINE RNA 3D models.

1031

1032 **Supplementary Table 4:** List of non-canonical interactions identified in SINE RNA 3D models.

1033

1034 **Supplementary Table 5:** Oligos used in (**a**) icSHAPE, (**b**) seCLIP, and (**c**) PARIS library preparation.

1035

1036 **Extended Data Figure 1. Effect of mouse natural antisense *Gadd45α* lncRNA on sense *Gadd45α* protein**
1037 **expression.** (**a**) Molecular features of mouse sense/antisense *Gadd45α* RNAs. Antisense (AS) *Gadd45α*
1038 contains a sequence complementary to intron 2, exon 3, and intron 3 of sense Gadd45α mRNA,two
1039 inverted SINEB2s, three LINE, and one SINE-Alu elements. (**b**) Western blot analysis of sense Gadd45α
1040 protein expression in HEK293T/17 cells co-transfected with sense/antisense *Gadd45α* for 24 h. Empty
1041 vector, negative control; plasmid transfection at + 0.7 pmol; ++ 0.8 pmol; +++ 0.9 pmol; ++++ 1.0 pmol; -
1042 not transfected. ACTB is the loading control.

1043

1044 **Extended Data Figure 2. qRT-PCR results for GFP mRNA (black bars) and SINEUP-GFP RNA (grey bars)**
1045 **expression.** Expression values are normalized to human GAPDH mRNA. Data were analyzed by using the
1046 $\Delta\Delta C_T$ method. RNA was extracted from (**a**) long SINEUPs (see description of terms in Figure 1c), with ΔAS
1047 *Uchl1* SINEB2 = 1 and from (**b**) MiniSINEUP-GFP (described in Figure 2b), with AS *Uchl1* = 1. Error bars
1048 are ± standard error mean for three biological replicates. AS, antisense.

1049

1050 **Extended Data Figure 3. Multiple sequence alignment and phylogenetic analysis of SINEB2 RNA**
1051 **sequences.** (**a**) ClustalW sequence alignment of mouse SINEB2 RNA sequences used as effector domain

1052 in long and mini SINEUPs. Gaps in the alignment are indicated by -. Bases are highlighted in four
1053 different colors. Black squares denote A and B boxes. Alignment is showing the sequence variation
1054 within members of the same family of SINE. (**b**) SINEB2 sequence alignment–based molecular
1055 phylogenetic analysis by the Maximum Likelihood method (Tamura 3-parameter model) to infer the
1056 evolutionary history. The tree with the highest log likelihood (–2415.7806) is shown. The tree is drawn
1057 to scale, with branch lengths measured in the number of substitutions per site. AS, antisense; B3 SINE2
1058 Rodentia RC, the consensus sequence for SINEB2 subfamily B3 from the Repbase database.

1059

1060 **Extended Data Figure 4. icSHAPE 2D structure models of mouse SINEB2 RNAs from natural antisense**
1061 **transcripts that are tested as effector domain in SINEUPs.** Regions that match the SINEUP motifs are
1062 marked by squares with their corresponding motif number. Blue square, AGG type motif; red square,
1063 UGG type motif. FC, GFP protein fold-change induced by the respective SINEB2. Nucleotide color
1064 indicates normalized icSHAPE reactivity score: red, ≥ 0.8; yellow, from above 0.5 to 0.8; white encircled
1065 with grey, ≤ 0.5; bases not encircled, no icSHAPE data available. AS, antisense.

1066

1067 **Extended Data Figure 5. SINEUP structure motifs in functionally inactive antisense (AS) *Gadd45α***
1068 **SINEB2-a deletion mutants.** Regions that match conserved SINEUP structure motifs are marked by
1069 different colors on the RNAfold-predicted structures of AS *Gadd45α* SINEB2-a mutants with deletions (Δ)
1070 at the positions shown in the figure.

1071

1072 **Extended Data Figure 6. Discovery of modified bases in SINEB2 RNA.** (**a**) An overview of SINEUP RNA
1073 modification analysis. SINEUP-GFP RNA (containing antisense Uchl1 SINEB2) transcribed in vitro was
1074 modified with 20% pseudouridine (Ψ), and a non-modified (0% Ψ) sample was used as control. In
1075 parallel, SINEUP-GFP and sense EGFP plasmids were co-transfected in HEK293T/17 cells, then SINEUP
1076 RNA that was transcribed in-cell was captured and enriched using specific biotinylated oligo probes and
1077 pull-down (twice) on streptavidin-coated magnetic beads. Purified samples of SINEUP-GFP RNA
1078 transcribed in-cell or transcribed and modified in vitro were sequenced by Nanopore direct RNA
1079 sequencing. Profiles of non-modified (0% Ψ) and modified SINEUP-GFP RNA transcribed in vitro were
1080 compared with that transcribed in-cell, and k-mers corresponding to the modified positions were
1081 identified in the in-cell transcripts. The red peaks in the graph mark the position of modified k-mers on
1082 the SINEUP transcript. The SINEB2 region is shaded in blue, and verified Ψ sites are marked. BD, binding
1083 domain. (**b**) Modified k-mers overlaid on 2D structure of antisense *Uchl1* SINEB2. k-mers are highlighted
1084 in yellow, Ψ sites are in red. SINEUP motifs are marked in black dashed squares with motifs 1–7 labelled
1085 as M1–M7.

1086

1087 **Extended Data Figure 7. Similarity of SINEUP motifs to rRNA 3D structure motifs.** (**a**) Hits found in the
1088 Characterization of Secondary Structure Motifs database for 3D motifs that match the SINEUP structure
1089 motifs. (**b**) The region similar to SINEUP motif 6 in helix 34 of human 18S rRNA is framed in black and
1090 marked with an arrow.

1091

1092 **Extended Data Figure 8. SINEUP-rRNA chimeric reads in PARIS sequencing data.** (**a**) SINEUP-GFP (of
1093 antisense *Uchl1* SINEB2) RNA chimeric reads with 18S, 5.8S, and 28S rRNA. Two biological replicates of
1094 the PARIS experiment are shown as separate tracks. SINEB2-rRNA interactions showed in Fig. 4 are
1095 indicated by respective colored arcs. (**b**) miniSINEUP-GFP FRAM RNA-rRNA chimeric reads. PARIS rep1
1096 and rep 2 are two biological replicates. The FRAM RNA-rRNA interactions displayed in Fig. 5 are marked
1097 by arcs in their respective colors.

1098

1099 **Extended Data Figure 9. SINE RNA predicted 3D structure models based on in-cell 2D structures of**
1100 **mouse SINEB2 RNAs.** SINEUP structure motifs are denoted from M1 to M7. AS: antisense.

1101

1102 **References**

1103

1104 1. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet* **30**, 121-3
1105 (2014).
1106 2. Novikova, I.V., Hennelly, S.P. & Sanbonmatsu, K.Y. Structural architecture of the human long
1107 non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* **40**, 5034-51 (2012).
1108 3. Somarowthu, S. et al. HOTAIR forms an intricate and modular secondary structure. *Mol Cell* **58**,
1109 353-61 (2015).
1110 4. Xue, Z. et al. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription
1111 Factor to Specify the Cardiovascular Lineage. *Mol Cell* **64**, 37-50 (2016).
1112 5. Guttman, M. & Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**,
1113 339-46 (2012).
1114 6. Johnson, R. & Guigo, R. The RIDL hypothesis: transposable elements as functional domains of
1115 long noncoding RNAs. *RNA* **20**, 959-76 (2014).
1116 7. Wilusz, J.E. et al. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A)
1117 tails. *Genes Dev* **26**, 2392-407 (2012).
1118 8. Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T. & Steitz, J.A. Formation of triple-helical
1119 structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad*
1120 *Sci U S A* **109**, 19202-7 (2012).
1121 9. Uroda, T. et al. Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53
1122 Pathway. *Mol Cell* **75**, 982-995 e9 (2019).
1123 10. Carrieri, C. et al. Long non-coding antisense RNA controls Uchl1 translation through an
1124 embedded SINEB2 repeat. *Nature* **491**, 454-7 (2012).
1125 11. Podbevsek, P. et al. Structural determinants of the SINE B2 element embedded in the long non-
1126 coding RNA activator of translation AS Uchl1. *Sci Rep* **8**, 3189 (2018).
1127 12. Ohyama, T. et al. An NMR-based approach reveals the core structure of the functional domain of
1128 SINEUP lncRNAs. *Nucleic Acids Res* **48**, 9346-9360 (2020).
1129 13. Kramerov, D.A. & Vassetzky, N.S. Short retroposons in eukaryotic genomes. *Int Rev Cytol* **247**,
1130 165-221 (2005).
1131 14. Ohshima, K. & Okada, N. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail.
1132 *Cytogenet Genome Res* **110**, 475-90 (2005).
1133 15. Su, M., Han, D., Boyd-Kirkup, J., Yu, X. & Han, J.J. Evolution of Alu elements toward enhancers.
1134 *Cell Rep* **7**, 376-385 (2014).

1135 16. Bourque, G. et al. Evolution of the mammalian transcription factor binding repertoire via
1136 transposable elements. *Genome Res* **18**, 1752-62 (2008).
1137 17. Sundaram, V. et al. Widespread contribution of transposable elements to the innovation of gene
1138 regulatory networks. *Genome Res* **24**, 1963-76 (2014).
1139 18. Roy-Engel, A.M. et al. Human retroelements may introduce intragenic polyadenylation signals.
1140 *Cytogenet Genome Res* **110**, 365-71 (2005).
1141 19. Policarpi, C. et al. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. *Cell Rep* **21**, 2879-
1142 2894 (2017).
1143 20. Zhang, X.O., Gingeras, T.R. & Weng, Z. Genome-wide analysis of polymerase III-transcribed Alu
1144 elements suggests cell-type-specific enhancer function. *Genome Res* **29**, 1402-1414 (2019).
1145 21. Faulkner, G.J. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet*
1146 **41**, 563-71 (2009).
1147 22. Fort, A. et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role
1148 for retrotransposons in pluripotency maintenance. *Nat Genet* **46**, 558-66 (2014).
1149 23. Lubelsky, Y. & Ulitsky, I. Sequences enriched in Alu repeats drive nuclear localization of long
1150 RNAs in human cells. *Nature* **555**, 107-111 (2018).
1151 24. Gong, C. & Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with
1152 3' UTRs via Alu elements. *Nature* **470**, 284-8 (2011).
1153 25. Mouse Genome Sequencing, C. et al. Initial sequencing and comparative analysis of the mouse
1154 genome. *Nature* **420**, 520-62 (2002).
1155 26. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome*
1156 *Res* **110**, 462-7 (2005).
1157 27. Bao, W., Kojima, K.K. & Kohany, O. Repbase Update, a database of repetitive elements in
1158 eukaryotic genomes. *Mob DNA* **6**, 11 (2015).
1159 28. Karijolich, J., Zhao, Y., Alla, R. & Glaunsinger, B. Genome-wide mapping of infection-induced
1160 SINE RNAs reveals a role in selective mRNA export. *Nucleic Acids Res* **45**, 6194-6208 (2017).
1161 29. Zucchelli, S. et al. SINEUPs are modular antisense long non-coding RNAs that increase synthesis
1162 of target proteins in cells. *Front Cell Neurosci* **9**, 174 (2015).
1163 30. Takahashi, H. et al. Identification of functional features of synthetic SINEUPs, antisense lncRNAs
1164 that specifically enhance protein translation. *PLoS One* **13**, e0183229 (2018).
1165 31. Schein, A., Zucchelli, S., Kauppinen, S., Gustincich, S. & Carninci, P. Identification of antisense
1166 long noncoding RNAs that function as SINEUPs in human cells. *Sci Rep* **6**, 33605 (2016).
1167 32. Sun, L. et al. RNA structure maps across mammalian cellular compartments. *Nat Struct Mol Biol*
1168 **26**, 322-330 (2019).
1169 33. Kierzek, E. et al. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic*
1170 *Acids Res* **42**, 3492-501 (2014).
1171 34. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated
1172 pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).
1173 35. Lin, T.Y., Mehta, R. & Glatt, S. Pseudouridines in RNAs: switching atoms means shifting
1174 paradigms. *FEBS Lett* **595**, 2310-2322 (2021).
1175 36. Rinn, J.L. et al. Functional demarcation of active and silent chromatin domains in human HOX
1176 loci by noncoding RNAs. *Cell* **129**, 1311-23 (2007).
1177 37. Tsai, M.C. et al. Long noncoding RNA as modular scaffold of histone modification complexes.
1178 *Science* **329**, 689-93 (2010).
1179 38. Toki, N. et al. SINEUP long non-coding RNA acts via PTBP1 and HNRNPK to promote translational
1180 initiation assemblies. *Nucleic Acids Res* **48**, 11626-11644 (2020).
1181 39. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B.J. Global Mapping of Human RNA-RNA
1182 Interactions. *Mol Cell* **62**, 618-26 (2016).

1183 40. Zhang, M. et al. Optimized photochemistry enables efficient analysis of dynamic RNA
1184 structuromes and interactomes in genetic and infectious diseases. *Nat Commun* **12**, 2344
1185 (2021).
1186 41. Meng, Z., Jackson, N.L., Shcherbakov, O.D., Choi, H. & Blume, S.W. The human IGF1R IRES likely
1187 operates through a Shine-Dalgarno-like interaction with the G961 loop (E-site) of the 18S rRNA
1188 and is kinetically modulated by a naturally polymorphic polyU loop. *J Cell Biochem* **110**, 531-44
1189 (2010).
1190 42. Chen, C.K. et al. Structured elements drive extensive circular RNA translation. *Mol Cell* **81**, 4300-
1191 4318 e13 (2021).
1192 43. Reisser, S., Zucchelli, S., Gustincich, S. & Bussi, G. Conformational ensembles of an RNA hairpin
1193 using molecular dynamics and sparse NMR data. *Nucleic Acids Res* **48**, 1164-1174 (2020).
1194 44. Spitale, R.C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**,
1195 486-90 (2015).
1196 45. Van Nostrand, E.L. et al. Robust, Cost-Effective Profiling of RNA Binding Protein Targets with
1197 Single-end Enhanced Crosslinking and Immunoprecipitation (seCLIP). *Methods Mol Biol* **1648**,
1198 177-200 (2017).
1199 46. Kalb, V.F., Glasser, S., King, D. & Lingrel, J.B. A cluster of repetitive elements within a 700 base
1200 pair region in the mouse genome. *Nucleic Acids Res* **11**, 2177-84 (1983).
1201 47. Lunyak, V.V. & Atallah, M. Genomic relationship between SINE retrotransposons, Pol III-Pol II
1202 transcription, and chromatin organization: the journey from junk to jewel. *Biochem Cell Biol* **89**,
1203 495-504 (2011).
1204 48. Flynn, R.A. et al. Transcriptome-wide interrogation of RNA secondary structure in living cells
1205 with icSHAPE. *Nat Protoc* **11**, 273-90 (2016).
1206 49. Heyne, S., Will, S., Beckstette, M. & Backofen, R. Lightweight comparison of RNAs based on
1207 exact sequence-structure matches. *Bioinformatics* **25**, 2095-102 (2009).
1208 50. Matsuda, D. & Mauro, V.P. Base pairing between hepatitis C virus RNA and 18S rRNA is required
1209 for IRES-dependent translation initiation in vivo. *Proc Natl Acad Sci U S A* **111**, 15385-9 (2014).
1210 51. Vanegas, P.L. et al. RNA CoSSMos: Characterization of Secondary Structure Motifs--a searchable
1211 database of secondary structure motifs in RNA three-dimensional structures. *Nucleic Acids Res*
1212 **40**, D439-44 (2012).
1213 52. Spahn, C.M. et al. Cryo-EM visualization of a viral internal ribosome entry site bound to human
1214 ribosomes: the IRES functions as an RNA-based translation factor. *Cell* **118**, 465-75 (2004).
1215 53. Lu, Z., Gong, J. & Zhang, Q.C. PARIS: Psoralen Analysis of RNA Interactions and Structures with
1216 High Throughput and Resolution. *Methods Mol Biol* **1649**, 59-84 (2018).
1217 54. Lu, Z. & Chang, H.Y. The RNA Base-Pairing Problem and Base-Pairing Solutions. *Cold Spring Harb*
1218 *Perspect Biol* **10**(2018).
1219 55. Vicens, Q. & Kieft, J.S. Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad*
1220 *Sci U S A* **119**, e2112677119 (2022).
1221 56. Vassetzky, N.S. & Kramerov, D.A. SINEBase: a database and tool for SINE analysis. *Nucleic Acids*
1222 *Res* **41**, D83-9 (2013).
1223 57. Kramerov, D.A. & Vassetzky, N.S. Origin and evolution of SINEs in eukaryotic genomes. *Heredity*
1224 *(Edinb)* **107**, 487-95 (2011).
1225 58. Davis, D.R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23**, 5020-6 (1995).
1226 59. Evans, J.R. et al. Members of the poly (rC) binding protein family stimulate the activity of the c-
1227 myc internal ribosome entry segment in vitro and in vivo. *Oncogene* **22**, 8012-20 (2003).
1228 60. Bou-Nader, C. & Zhang, J. Structural Insights into RNA Dimerization: Motifs, Interfaces and
1229 Functions. *Molecules* **25**(2020).

1230 61. Spahn, C.M. et al. Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-
1231 ribosome and subunit-subunit interactions. *Cell* **107**, 373-86 (2001).
1232 62. Simonetti, A., Guca, E., Bochler, A., Kuhn, L. & Hashem, Y. Structural Insights into the
1233 Mammalian Late-Stage Initiation Complexes. *Cell Rep* **31**, 107497 (2020).
1234 63. Unbehaun, A. et al. Position of eukaryotic initiation factor eIF5B on the 80S ribosome mapped
1235 by directed hydroxyl radical probing. *EMBO J* **26**, 3109-23 (2007).
1236 64. Armache, J.P. et al. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome
1237 at 5.5-A resolution. *Proc Natl Acad Sci U S A* **107**, 19748-53 (2010).
1238 65. Heuer, A. et al. Cryo-EM structure of a late pre-40S ribosomal subunit from Saccharomyces
1239 cerevisiae. *Elife* **6**(2017).
1240 66. Tamura, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
1241 evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9 (2011).
1242 67. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
1243 68. Zarringhalam, K., Meyer, M.M., Dotu, I., Chuang, J.H. & Clote, P. Integrating chemical
1244 footprinting data into RNA secondary structure prediction. *PLoS One* **7**, e45160 (2012).
1245 69. Johnson, P.Z., Kasprzak, W.K., Shapiro, B.A. & Simon, A.E. RNA2Drawer: geometrically strict
1246 drawing of nucleic acid structures with graphical structure editing and highlighting of
1247 complementary subsequences. *RNA Biol* **16**, 1667-1671 (2019).
1248 70. Pandya-Jones, A. & Black, D.L. Co-transcriptional splicing of constitutive and alternative exons.
1249 *RNA* **15**, 1896-908 (2009).
1250 71. Brugiolo, M., Botti, V., Liu, N., Muller-McNicoll, M. & Neugebauer, K.M. Fractionation iCLIP
1251 detects persistent SR protein binding to conserved, retained introns in chromatin, nucleoplasm
1252 and cytoplasm. *Nucleic Acids Res* **45**, 10452-10465 (2017).
1253 72. Toki, N., Takahashi, H., Zucchelli, S., Gustincich, S. & Carninci, P. Synthetic in vitro transcribed
1254 lncRNAs (SINEUPs) with chemical modifications enhance target mRNA translation. *FEBS Lett* **594**,
1255 4357-4369 (2020).
1256 73. Mandal, P.K. & Rossi, D.J. Reprogramming human fibroblasts to pluripotency using modified
1257 mRNA. *Nat Protoc* **8**, 568-82 (2013).
1258 74. Leger, A. et al. RNA modifications detection by comparative Nanopore direct RNA sequencing.
1259 *Nat Commun* **12**, 7198 (2021).
1260 75. Zhang, S., Cheng, Y., Guo, P. & Chen, S.J. VfoldMCPX: predicting multistrand RNA complexes.
1261 *RNA* **28**, 596-608 (2022).
1262 76. Biesiada, M., Purzycka, K.J., Szachniuk, M., Blazewicz, J. & Adamiak, R.W. Automated RNA 3D
1263 Structure Prediction with RNAComposer. *Methods Mol Biol* **1490**, 199-215 (2016).
1264 77. Sehnal, D. et al. Mol* Viewer: modern web app for 3D visualization and analysis of large
1265 biomolecular structures. *Nucleic Acids Res* **49**, W431-W437 (2021).
1266 78. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).
1267 79. Zok, T. et al. RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res*
1268 **46**, W30-W35 (2018).

1269

Figure 1

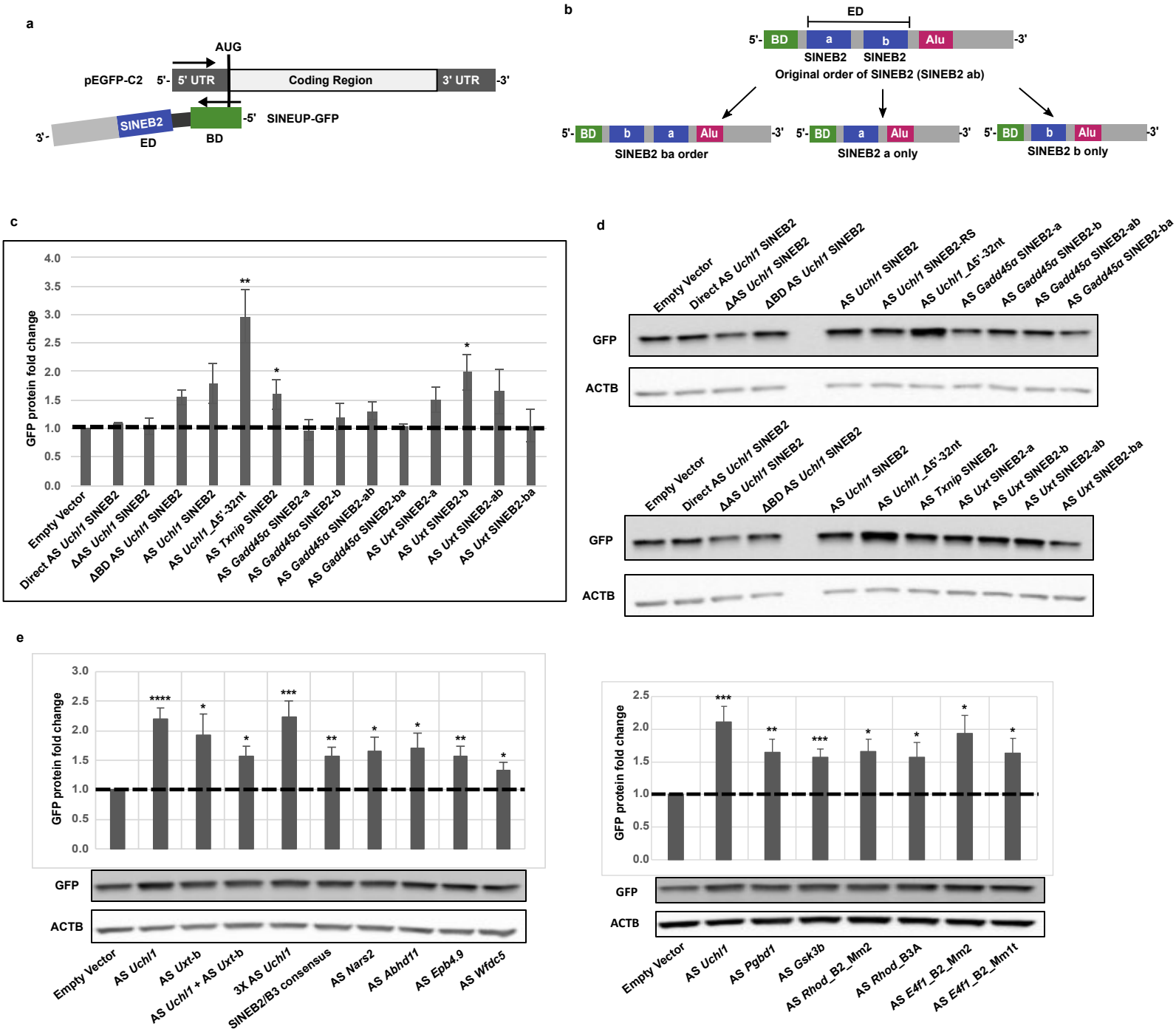

a
AUG
pEGFP-C2
5'-
5' UTR
Coding Region
3' UTR
-3'
3'-
SINEB2
ED
BD
-5'
SINEUP-GFP
b
ED
5'-
BD
a
b
Alu
-3'
SINEB2
SINEB2
Original order of SINEB2 (SINEB2 ab)
SINEB2 ba order
SINEB2 a only
SINEB2 b only
c
GFP protein fold change
Empty Vector
Direct AS Uchl1 SINEB2
ΔAS Uchl1 SINEB2
ΔBD AS Uchl1 SINEB2
AS Uchl1 SINEB2
AS Uchl1_Δ5'-32nt
AS Txnip SINEB2
AS Gadd45α SINEB2-a
AS Gadd45α SINEB2-b
AS Gadd45α SINEB2-ab
AS Gadd45α SINEB2-ba
AS Uxt SINEB2-a
AS Uxt SINEB2-b
AS Uxt SINEB2-ab
AS Uxt SINEB2-ba
d
AS Uchl1 SINEB2-RS
GFP
ACTB
e
AS Uxt-b
AS Uxt-b + AS Uxt-b
3X AS Uchl1
SINEB2/B3 consensus
AS Nars2
AS Abhd11
AS Epb4.9
AS Wfdc5
AS Pgbd1
AS Gsk3b
AS Rhod_B2_Mm2
AS Rhod_B3A
AS E4f1_B2_Mm2
AS E4f1_B2_Mm1t

Figure 2

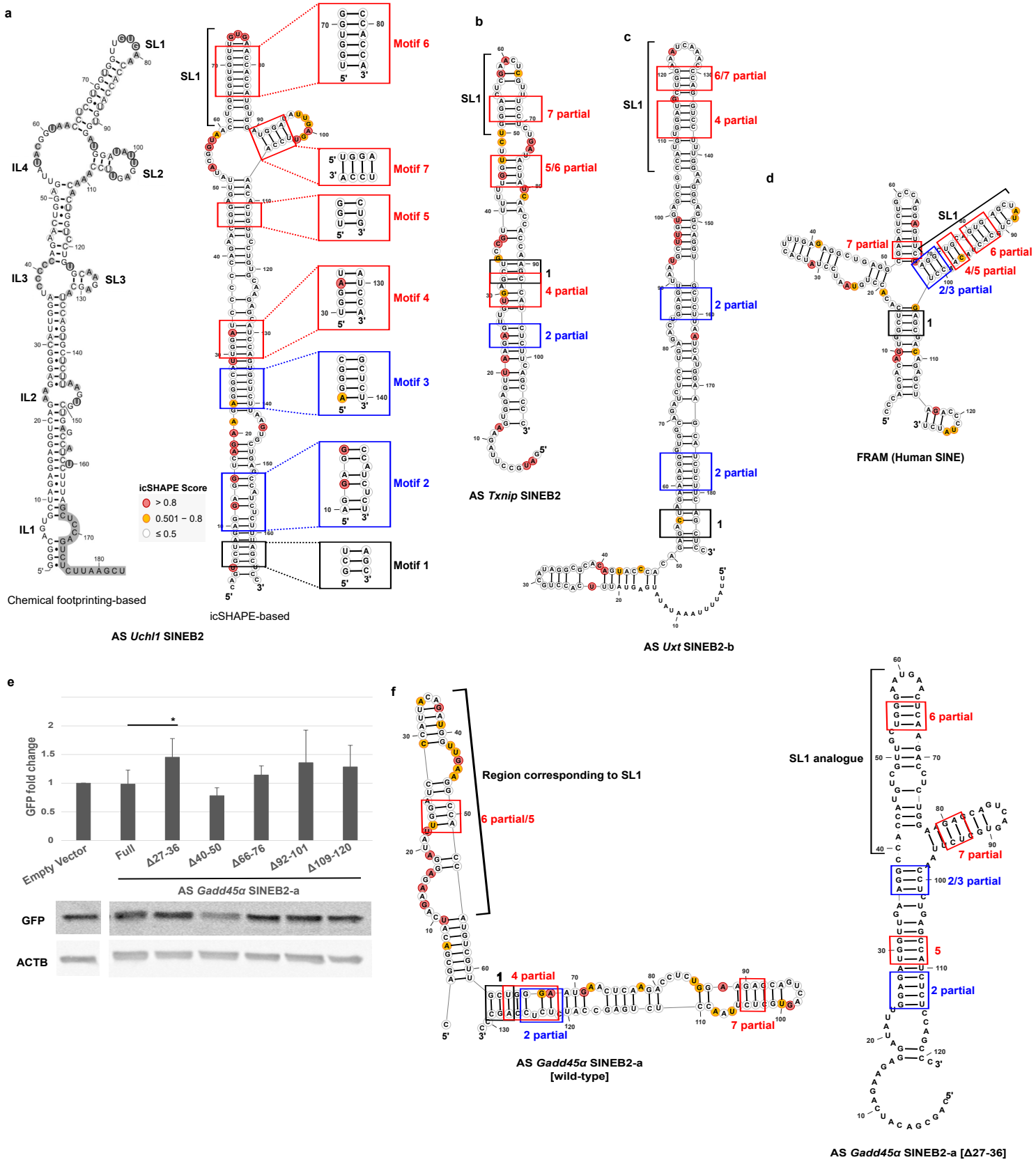

a
SL1
SL2
SL3
IL1
IL2
IL4
Chemical footprinting-based
icSHAPE Score
> 0.8
0.501 – 0.8
≤ 0.5
Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 6
Motif 7
icSHAPE-based
AS Uchl1 SINEB2
b
SL1
7 partial
5/6 partial
4 partial
2 partial
AS Txnip SINEB2
c
SL1
6/7 partial
4 partial
2 partial
2 partial
AS Uxt SINEB2-b
d
SL1
7 partial
6 partial
4/5 partial
2/3 partial
FRAM (Human SINE)
e
GFP fold change
Empty Vector
Full
Δ27-36
Δ40-50
Δ66-76
Δ92-101
Δ109-120
AS Gadd45α SINEB2-a
GFP
ACTB
f
Region corresponding to SL1
6 partial/5
4 partial
2 partial
7 partial
AS Gadd45α SINEB2-a [wild-type]
SL1 analogue
6 partial
7 partial
2/3 partial
2 partial
AS Gadd45α SINEB2-a [Δ27-36]

Figure 3

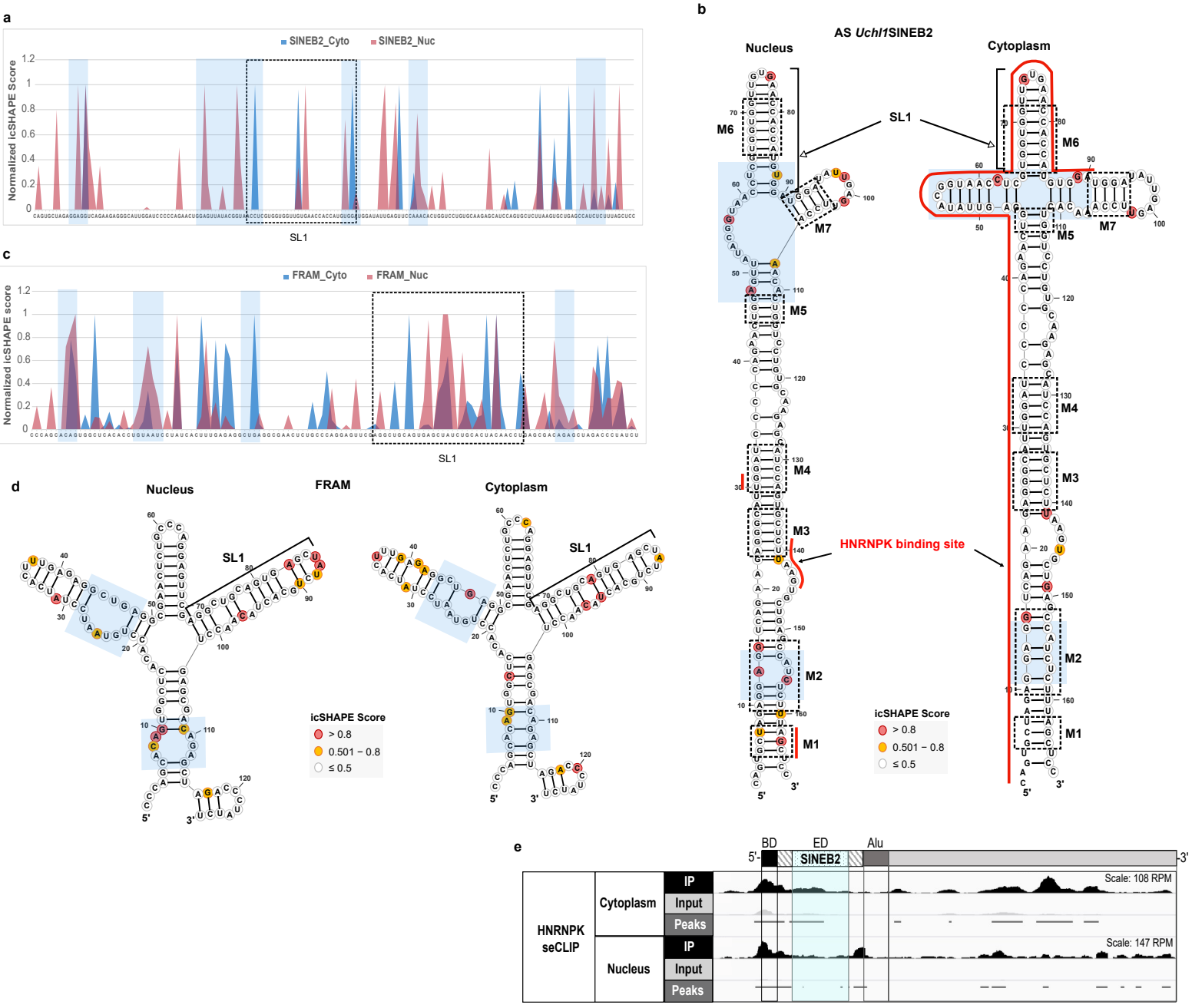

a
SINEB2_Cyto
SINEB2_Nuc
Normalized icSHAPE Score
SL1
b
AS Uchl1SINEB2
Nucleus
Cytoplasm
SL1
M6
M7
M5
M4
M3
M2
M1
HNRNPK binding site
icSHAPE Score
> 0.8
0.501 – 0.8
≤ 0.5
5'
3'
c
FRAM_Cyto
FRAM_Nuc
Normalized icSHAPE score
SL1
d
Nucleus
FRAM
Cytoplasm
SL1
icSHAPE Score
> 0.8
0.501 – 0.8
≤ 0.5
5'
3'
e
BD
ED
Alu
SINEB2
5'
3'
HNRNPK seCLIP
Cytoplasm
Nucleus
IP
Input
Peaks
Scale: 108 RPM
Scale: 147 RPM

Figure 4

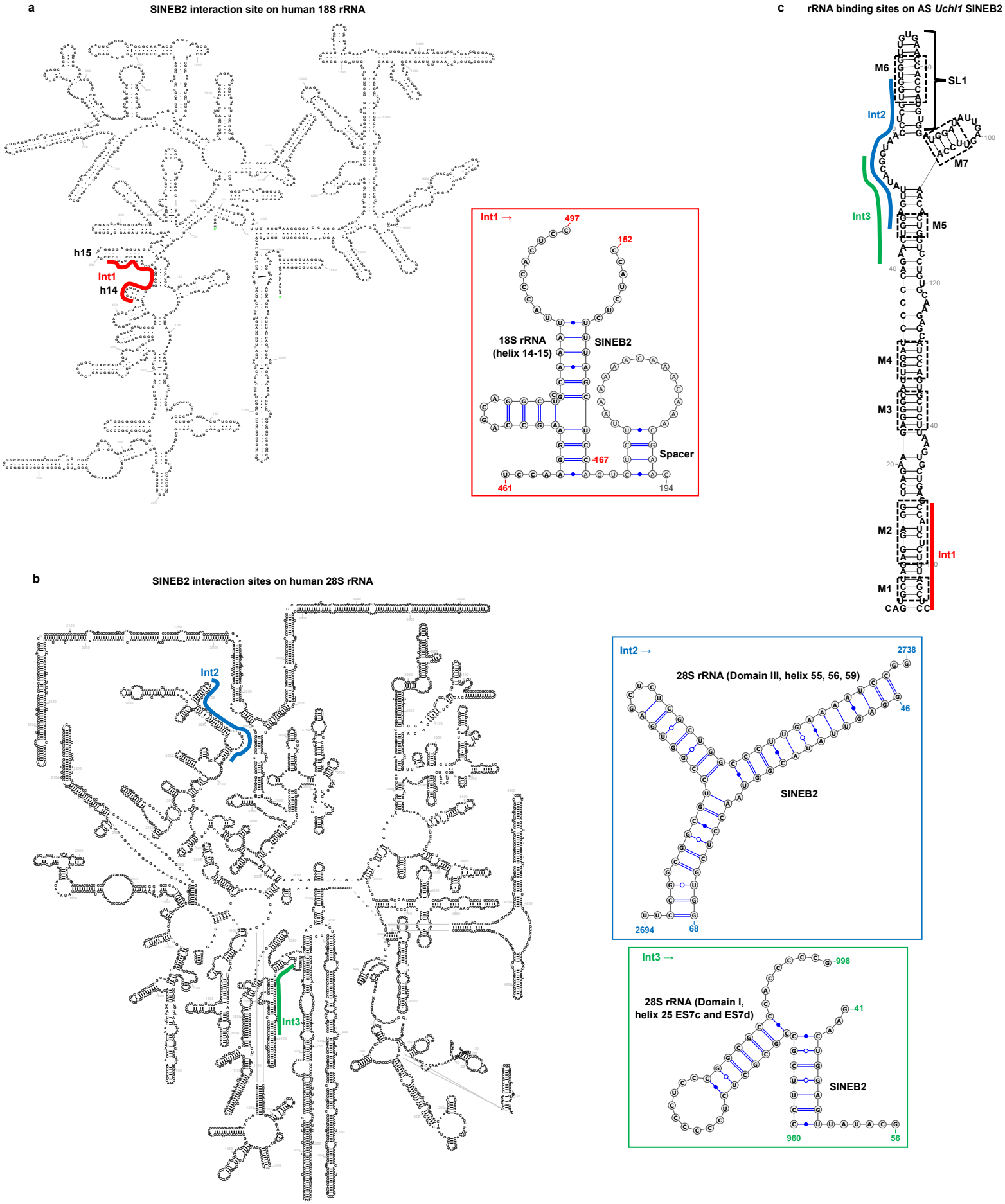

a
SINEB2 interaction site on human 18S rRNA
h15
Int1
h14
Int1 →
497
152
18S rRNA
(helix 14-15)
SINEB2
Spacer
-167
461
194
b
SINEB2 interaction sites on human 28S rRNA
Int2
Int3
Int2 →
2738
28S rRNA (Domain III, helix 55, 56, 59)
46
SINEB2
2694
68
Int3 →
-998
28S rRNA (Domain I,
helix 25 ES7c and ES7d)
-41
SINEB2
960
56
c
rRNA binding sites on AS Uchl1 SINEB2
M6
SL1
Int2
M7
Int3
M5
M4
M3
M2
Int1
M1

Figure 5 (Part 1)

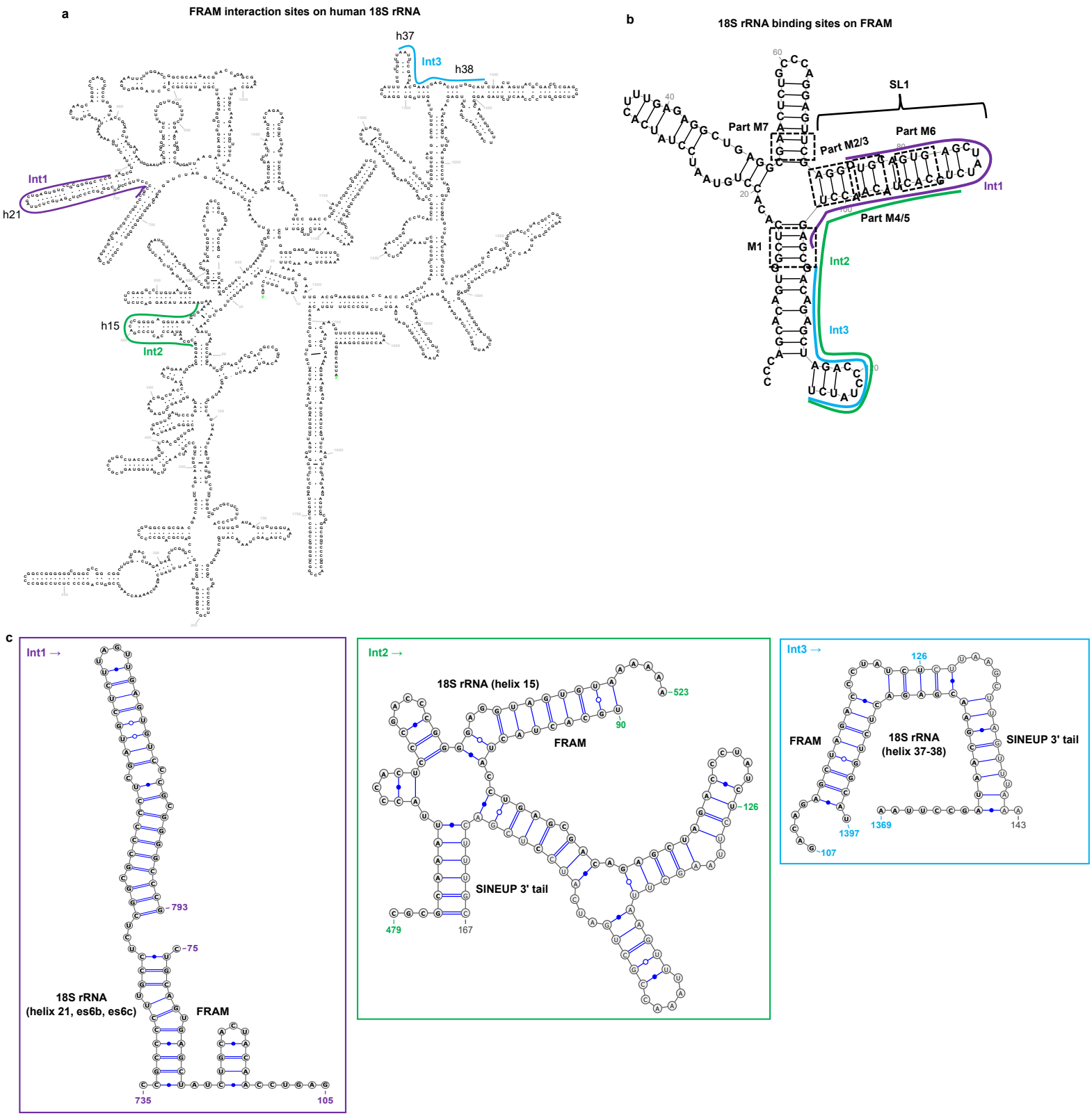


Figure 5 (Part 2)

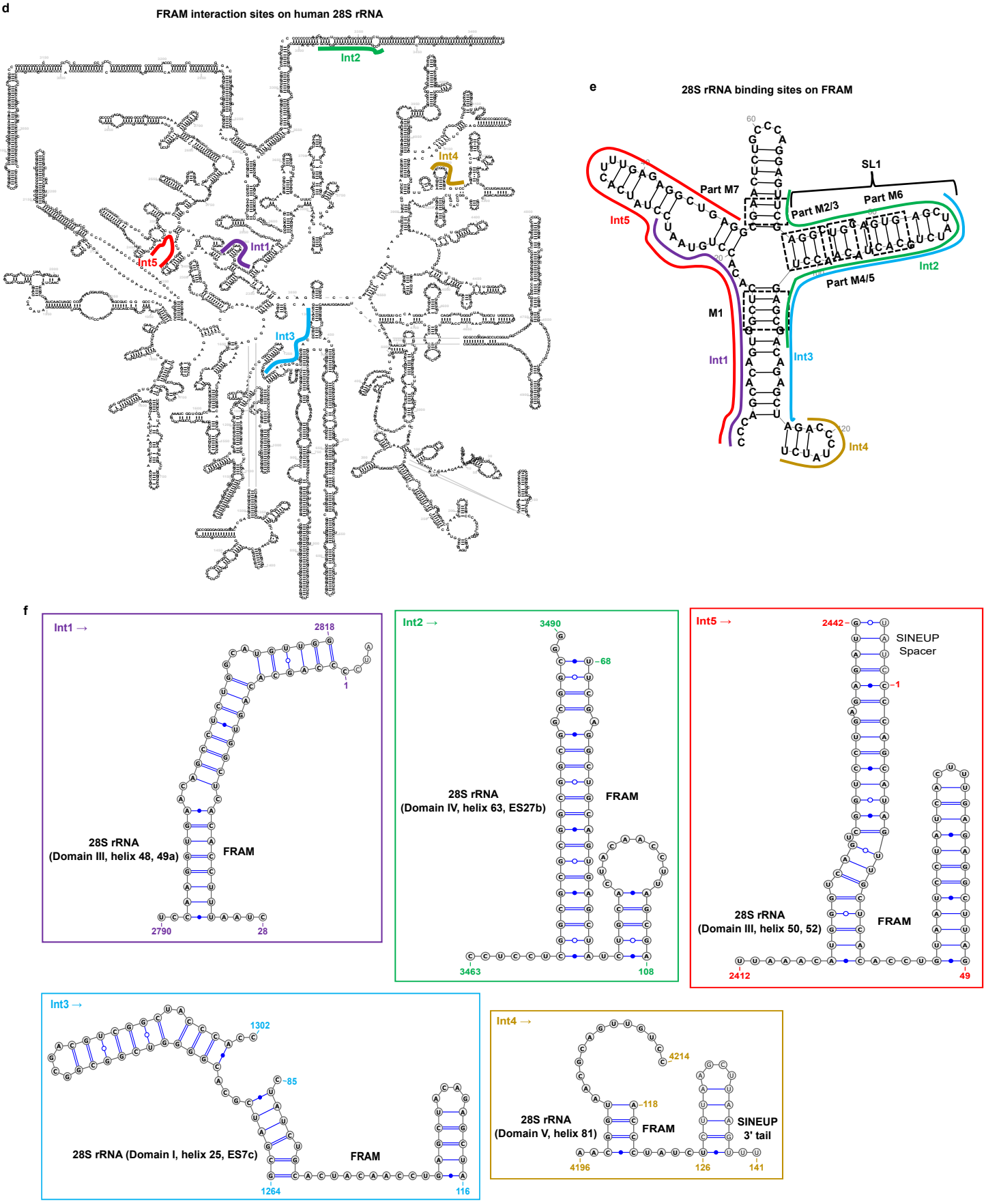
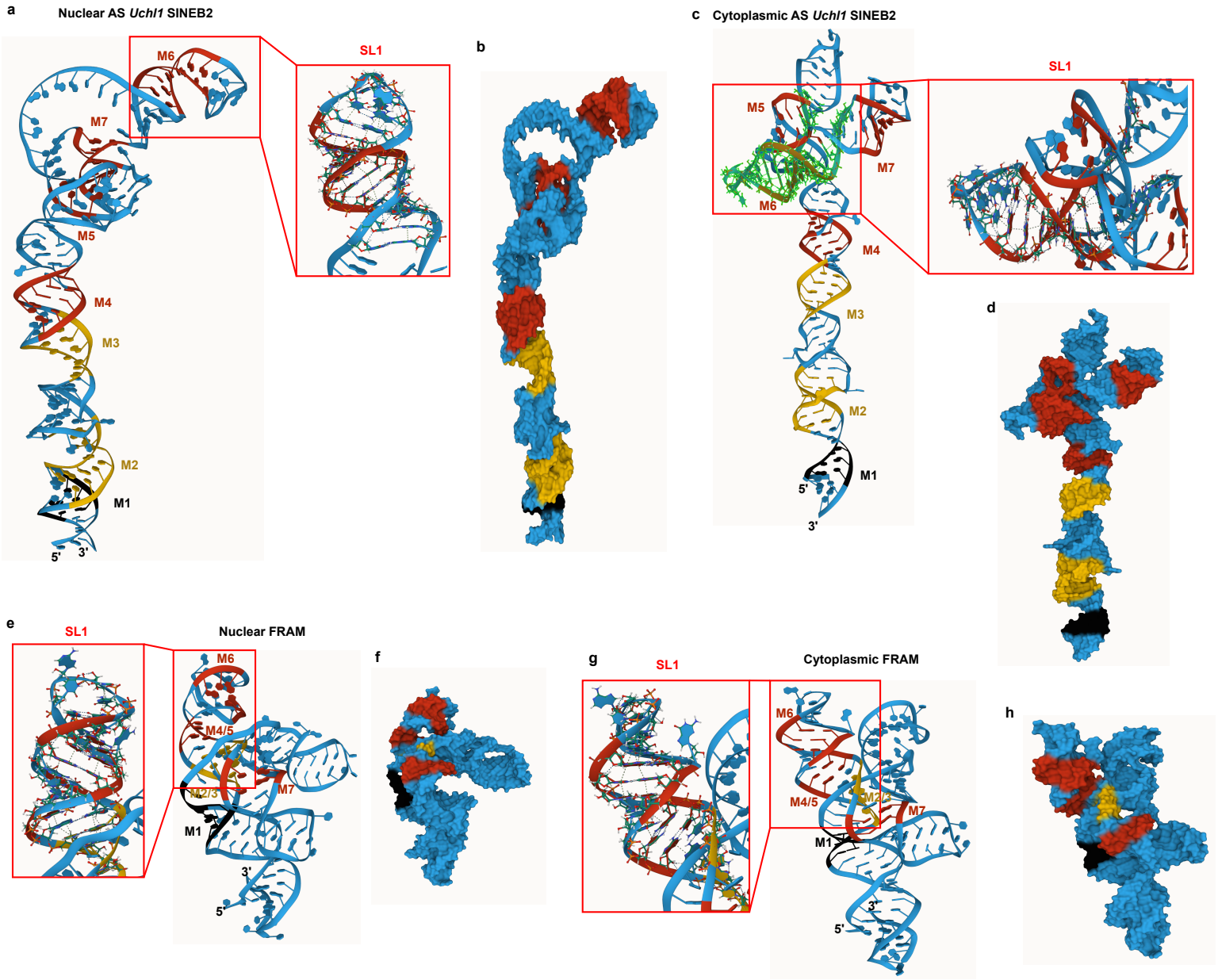
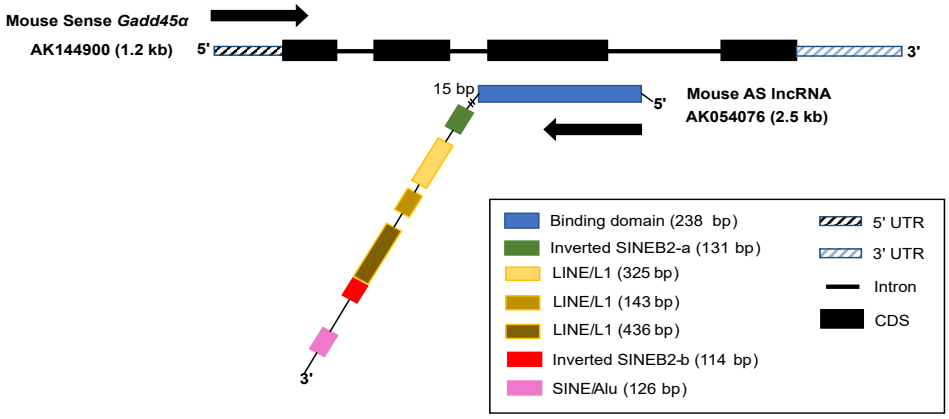


Figure 6

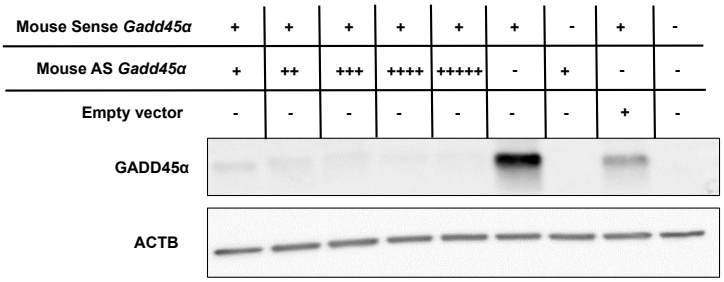

a
Nuclear AS *Uchl1* SINEB2
M6
SL1
M7
M5
M4
M3
M2
M1
5'
3'
b
c
Cytoplasmic AS *Uchl1* SINEB2
SL1
M5
M7
M6
M4
M3
M2
M1
5'
3'
d
e
SL1
Nuclear FRAM
M6
M4/5
M2/3
M7
M1
3'
5'
f
g
SL1
Cytoplasmic FRAM
M6
M4/5
M2/3
M7
M1
3'
5'
h

Extended Data Figure 1

a



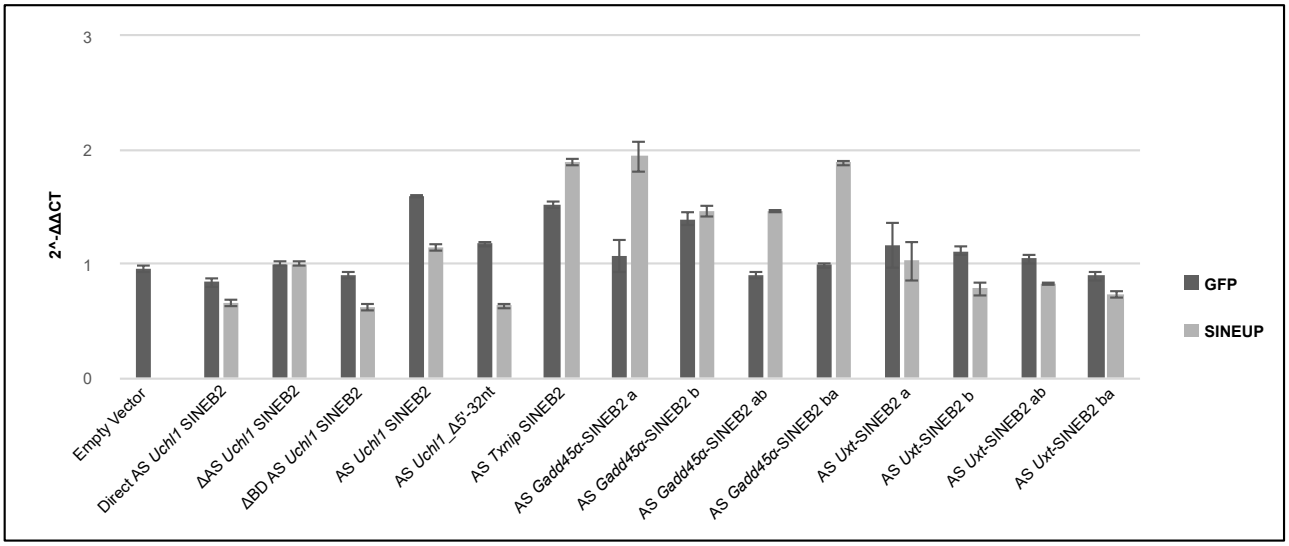
b



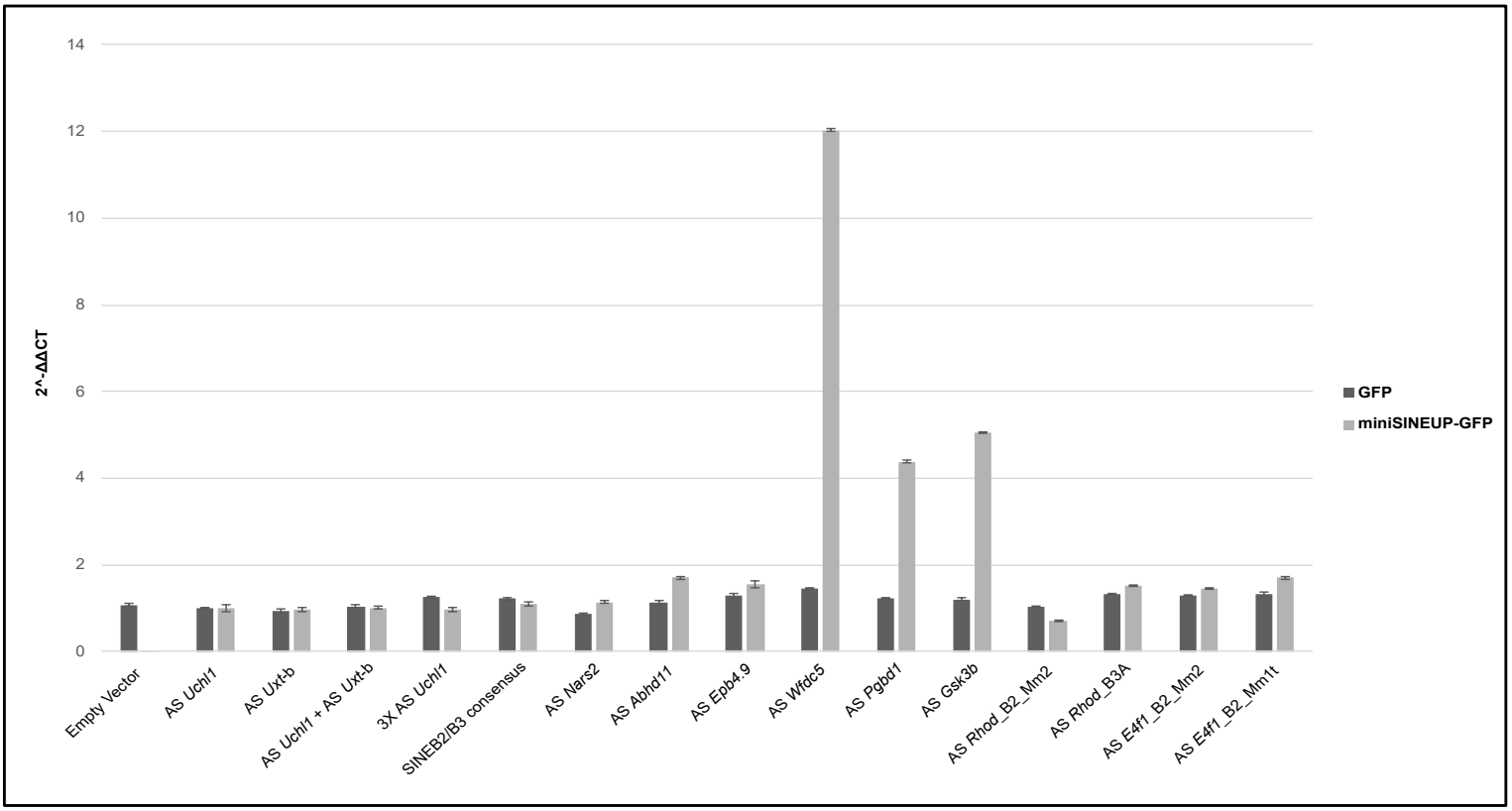
Mouse Sense *Gadd45α*	+	+	+	+	+	+	-	+	-
Mouse AS *Gadd45α*	+	++	+++	++++	+++++	-	+	-	-
Empty vector	-	-	-	-	-	-	-	+	-

Extended Data Figure 2

a

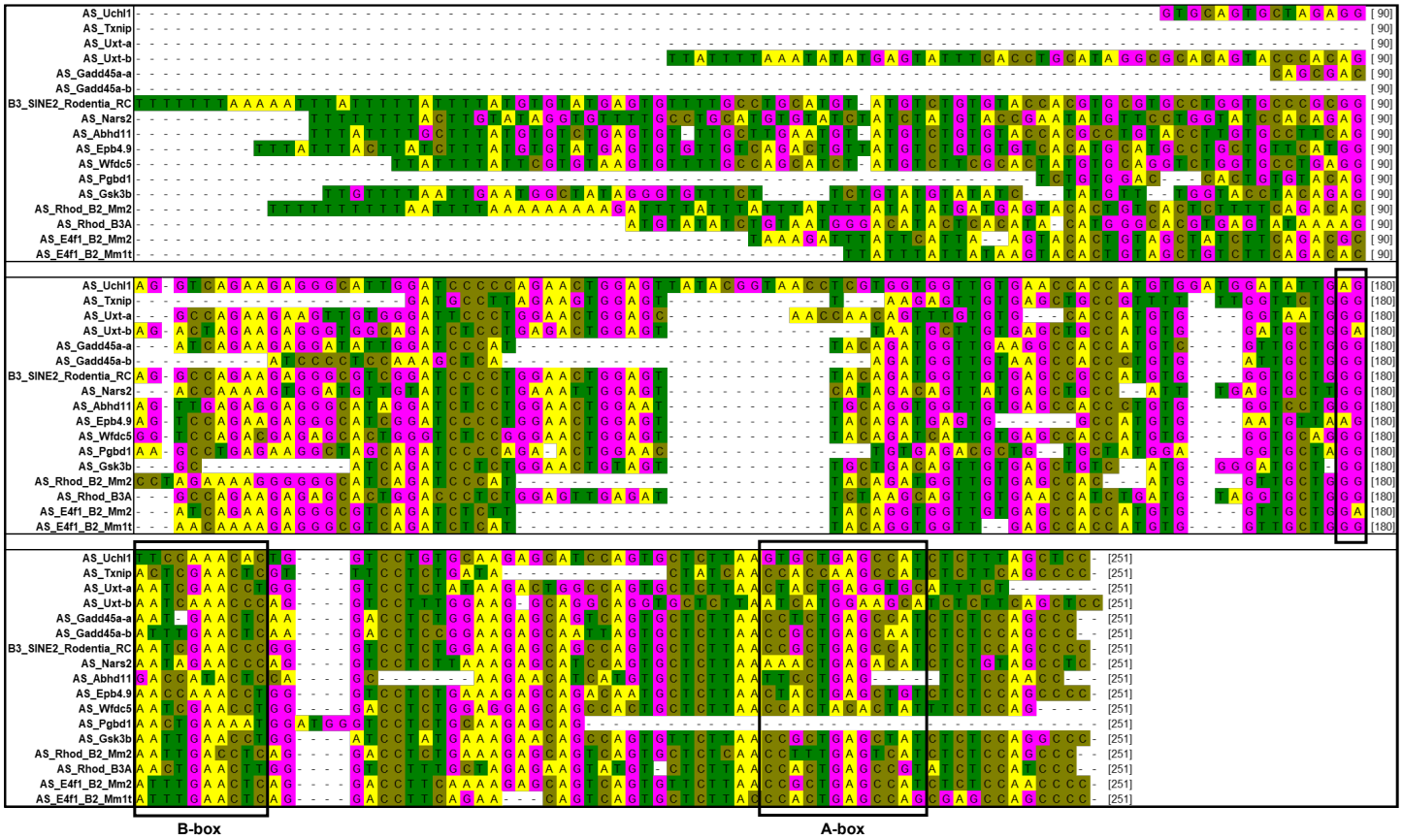


b

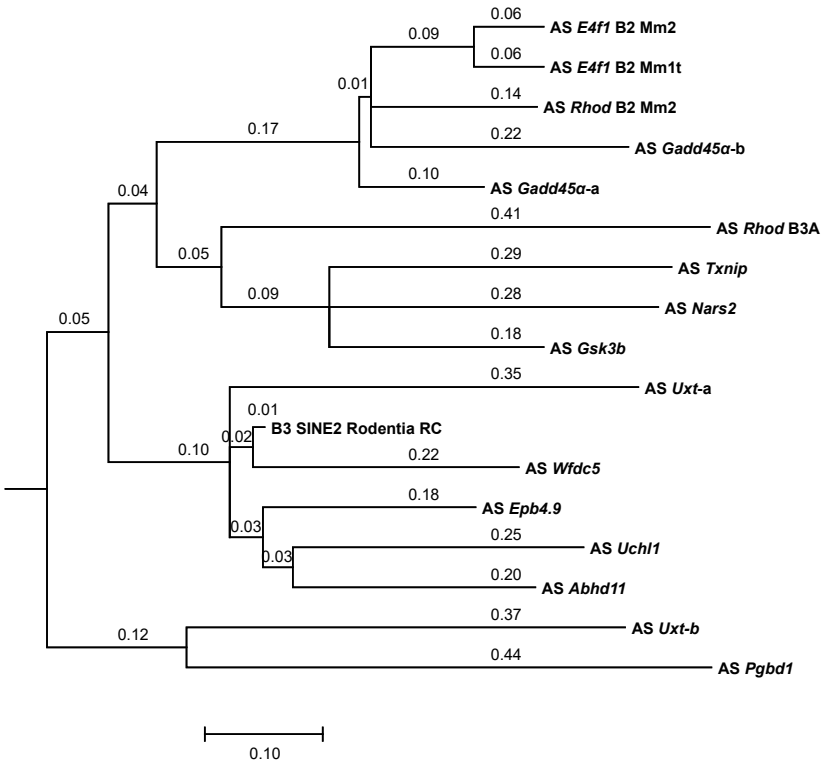


Extended Data Figure 3

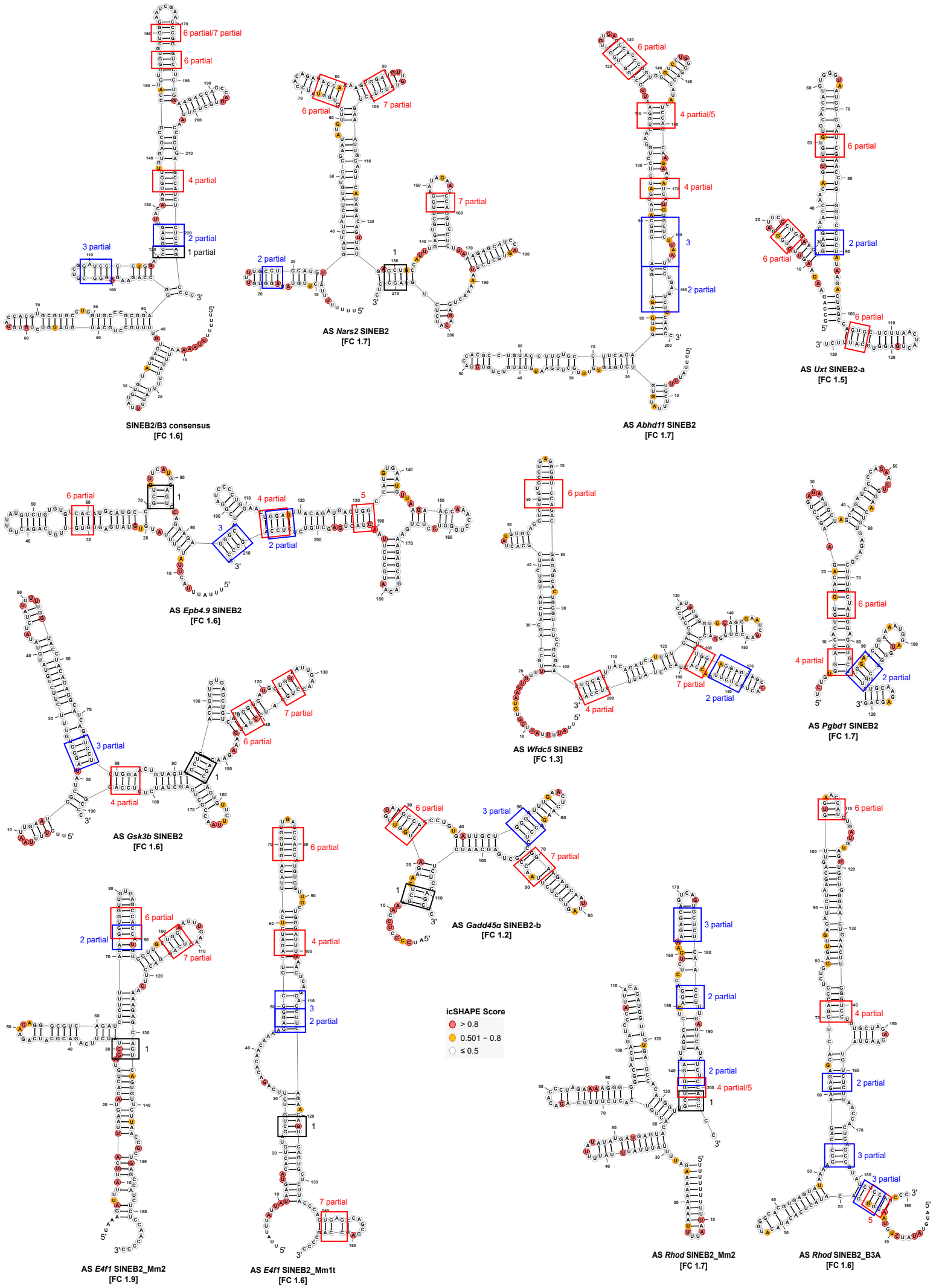
a



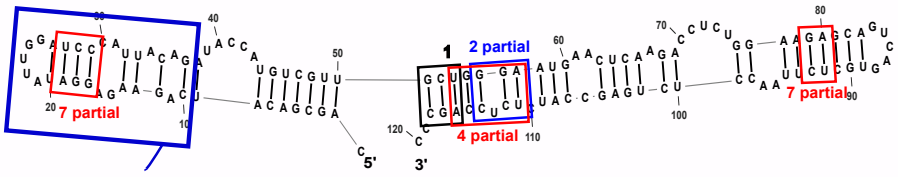
b



Extended Data Figure 4

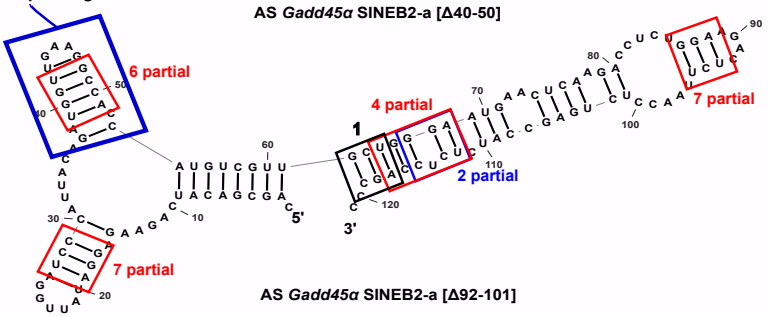


Extended Data Figure 5

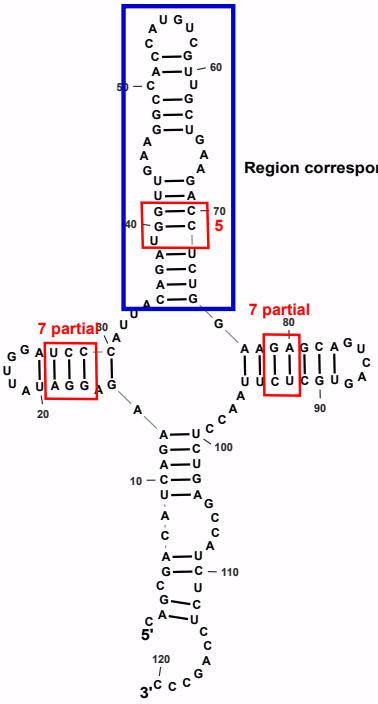


Region corresponding to SL1

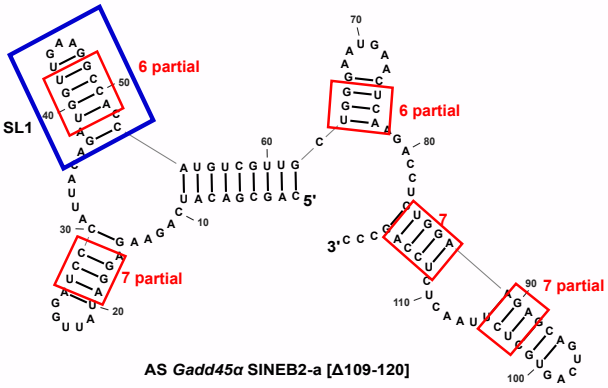
AS *Gadd45α* SINEB2-a [Δ40-50]



AS *Gadd45α* SINEB2-a [Δ92-101]



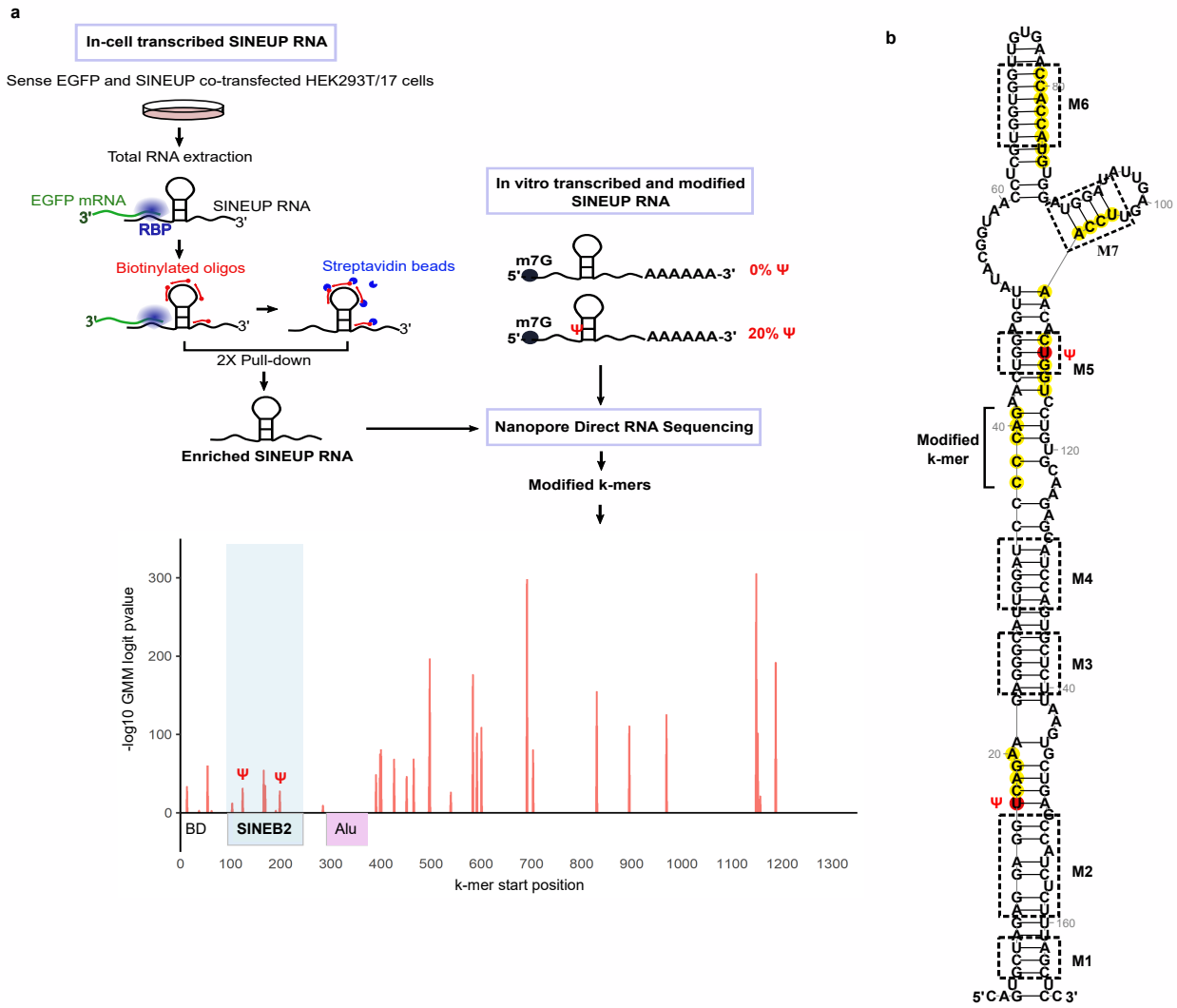
AS *Gadd45α* SINEB2-a [Δ66-76]



Region corresponding to SL1

AS *Gadd45α* SINEB2-a [Δ109-120]

Extended Data Figure 6

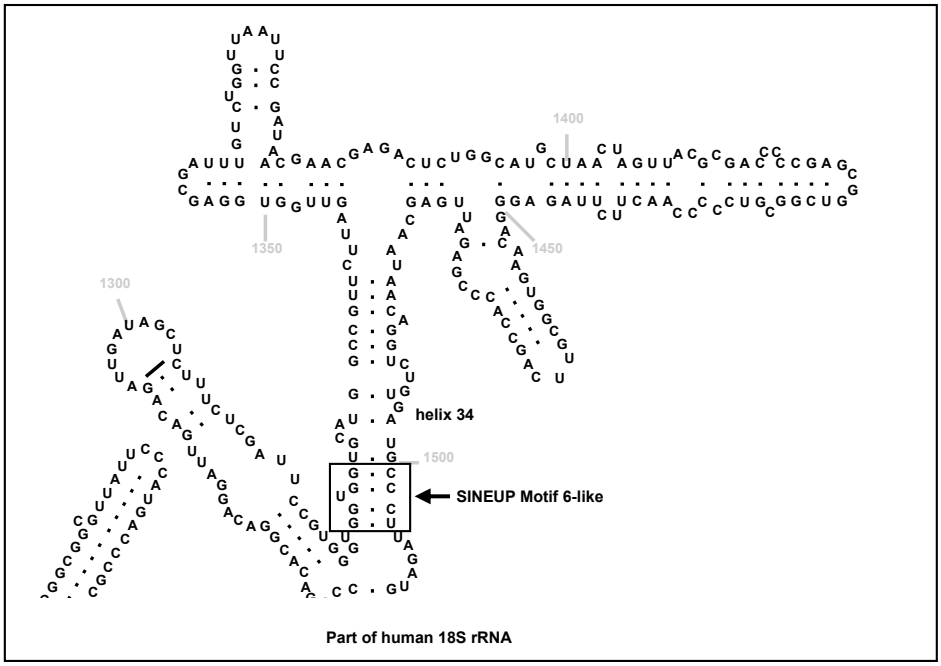


Extended Data Figure 7

a

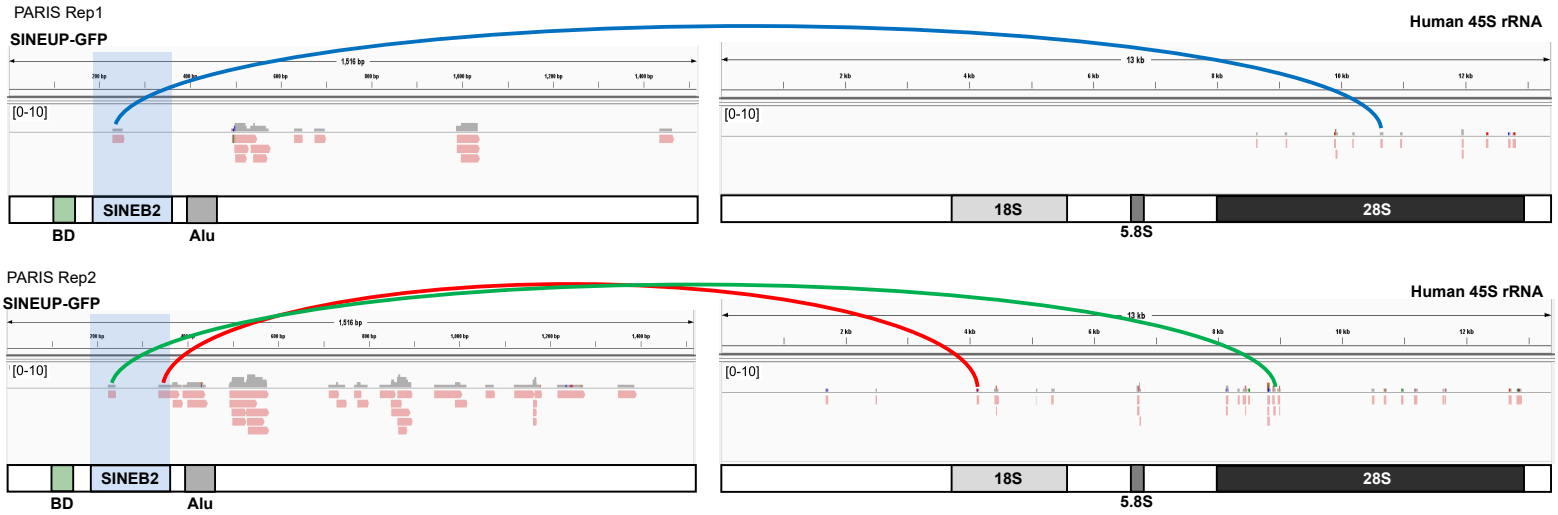
S. No.	SINEUP Motif	CoSSMos motif match	PDB ID	Positions	Molecule
1.	Motif 6	GGUGG UCCC	5A2Q	G1324-G1325-U1326-G1327-G1328 U1504-C1503-C1502-C1501	18S rRNA
2.	Motif 2	GGAGG CCCC	5JB3	G999-G1000-A1001-G1002-G1003 C1170-C1169-C1168-C1167	16S rRNA

b

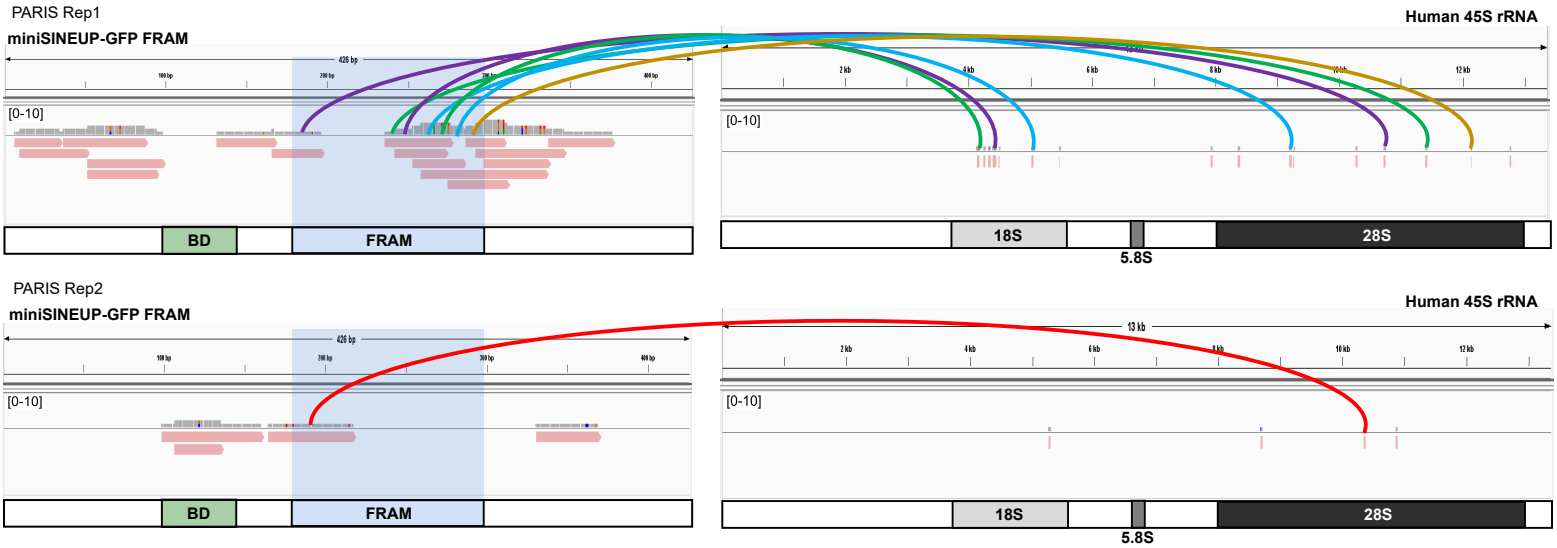


Extended Data Figure 8

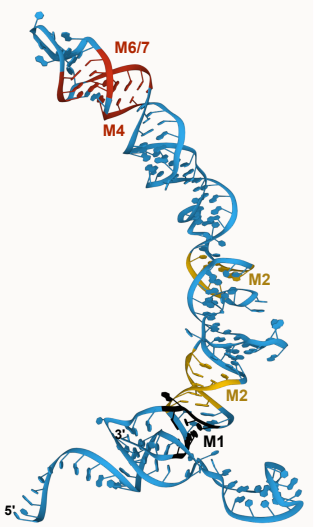
a



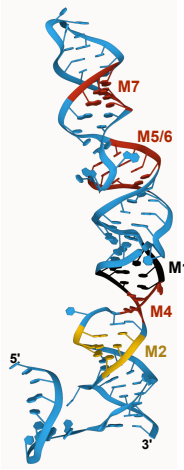
b



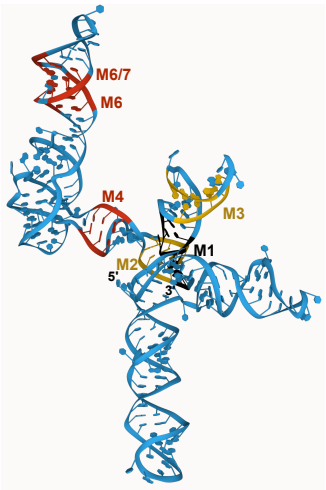
Extended Data Figure 9



AS *Uxt* SINEB2-b

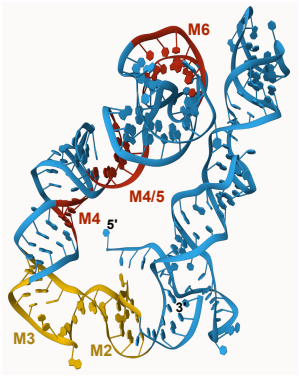


AS *Txnip* SINEB2

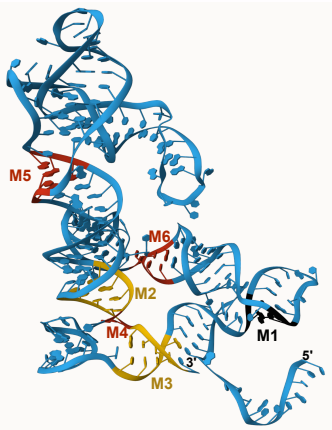


SINEB2/B3 consensus

AS *Nars2* SINEB2

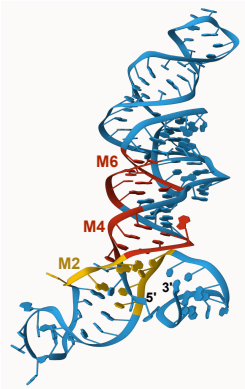


AS *Abhd11* SINEB2

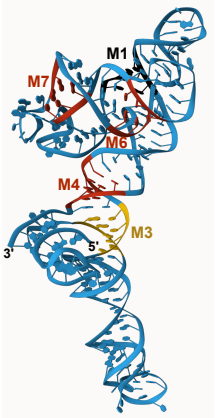


AS *Epb4.9* SINEB2

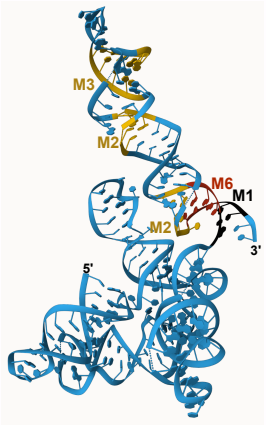
AS *Wfdc5* SINEB2



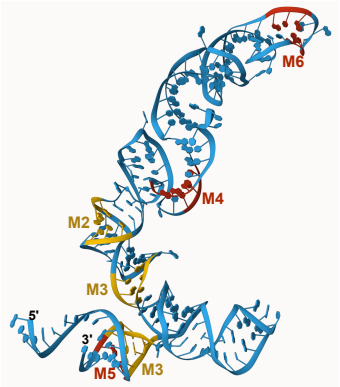
AS *Pgbd1* SINEB2



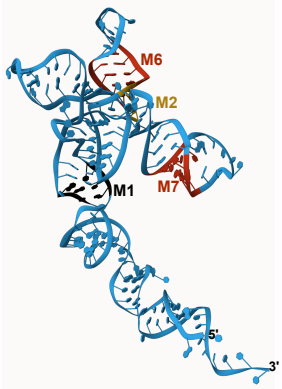
AS *Gsk3b* SINEB2



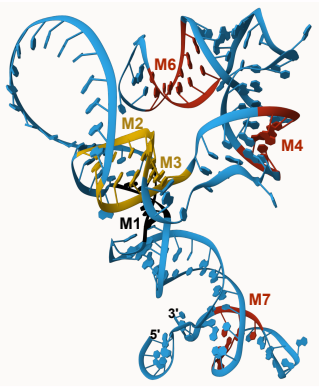
AS *Rhod* SINEB2_Mm2



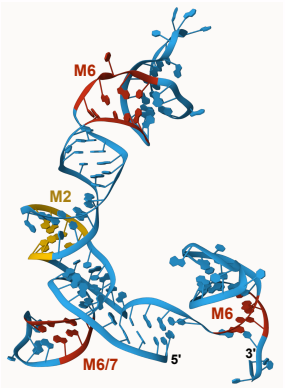
AS *Rhod* SINEB2_B3A



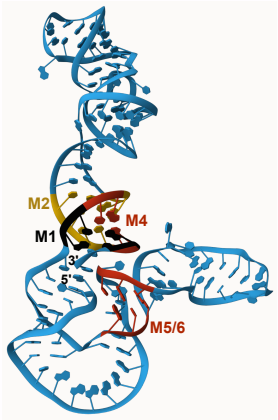
AS *E4f1* SINEB2_Mm2



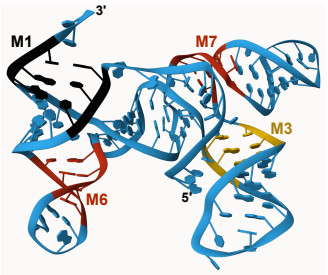
AS *E4f1* SINEB2_Mm1t



AS *Uxt* SINEB2-a



AS *Gadd45α* SINEB2-a



AS *Gadd45α* SINEB2-b