

An antisense RNA capable of modulating the expression of the tumor suppressor microRNA-34a

Jason T. Serviss^{1,*}, Felix Clemens Richter^{1,2}, Jimmy Van den Eynden³, Nathanael Andrews¹, Miranda Houtman^{1,4}, Mattias Vesterlund¹, Laura Schwarzmueller^{1,5}, Per Johnsson^{6,7}, Erik Larsson³, Dan Grandér¹ †, Katja Pokrovskaja Tamm¹

¹ Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden, SE-17177

² Kennedy Institute of Rheumatology, University of Oxford, Roosevelt Drive, Oxford OX3 7FY, UK

³ Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-405 30 Gothenburg, Sweden

⁴ Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Solna, Karolinska Institutet, Stockholm, Sweden

⁵ Laboratory for Experimental Oncology and Radiobiology (LEXOR), Center for Experimental Molecular Medicine (CEMM), Academic Medical Center, Amsterdam, The Netherlands

⁶ Ludwig Institute for Cancer Research, Stockholm, Sweden

¹ Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden

* Correspondence:

Jason T. Serviss
Department of Oncology and Pathology

Department of Oncology and Pathology Karolinska Institutet

Karolinska Institutet
Stockholm Sweden

Stockholm, Sweden
+46 8 517 75 43

jason.serviss@ki.se

jason.serviss@rl.ac.uk

Abstract

Abstract

The microRNA-34a is a well-studied tumor suppressor microRNA (miRNA) and a direct downstream target of TP53 with roles in several pathways associated with oncogenesis, such as proliferation, cellular growth, and differentiation. Due to its broad tumor suppressive activity, it is not surprising that *miR34a* expression is altered in a wide variety of solid tumors and hematological malignancies. However, the mechanisms by which *miR34a* is regulated in these cancers is largely unknown. In this study, we find that a long non-coding RNA transcribed antisense to the *miR34a* host gene, is critical for *miR34a* expression and mediation of its cellular functions in multiple types of human cancer. We name this long non-coding RNA *lncTAM34a*, and

42 characterize its ability to facilitate *miR34a* expression under different types of
43 cellular stress in both *TP53* deficient and wild type settings.

44

45 **Introduction**

46 In recent years advances in functional genomics have revolutionized our
47 understanding of the human genome. Evidence now points to the fact that
48 approximately 75% of the genome is transcribed but only ~1.2% of this is
49 responsible for encoding proteins^{1,2}. Of these recently identified elements,
50 long non-coding (lnc) RNAs are defined as transcripts exceeding 200 base
51 pairs (bp) in length with a lack of a functional open reading frame. Some
52 lncRNAs are dually classified as antisense (as) RNAs that are expressed from
53 the same locus as a sense transcript in the opposite orientation. Current
54 estimates using high-throughput transcriptome sequencing, indicate that up to
55 20-40% of the approximately 20,000 protein-coding genes exhibit antisense
56 transcription³⁻⁵.

57

58 Systematic large-scale studies have shown aberrant expression of asRNAs to
59 be associated with tumorigenesis⁶ and, although characterization of several of
60 these has identified asRNA-mediated regulation of multiple well known
61 tumorigenic factors^{7,8}, the vast majority of potential tumor-associated asRNAs
62 have not yet been characterized. The known mechanisms by which asRNAs
63 accomplish their regulatory functions are diverse, and include recruitment of
64 chromatin modifying factors^{8,9}, acting as microRNA (miRNA) sponges¹⁰, and
65 causing transcriptional interference¹¹.

66

67 Responses to cellular stress, e.g. DNA damage, sustained oncogene
68 expression, and nutrient deprivation, are all tightly controlled cellular pathways
69 that are almost universally dysregulated in cancer. Cellular signaling, in
70 response to these types of stresses, often converges on the transcription
71 factor TP53 that regulates transcription of coding and non-coding downstream
72 targets. One important non-coding target of TP53 is the tumor suppressor
73 miRNA known as *miR34a*¹². Upon TP53 activation *miR34a* expression is
74 increased allowing it to down-regulate target genes involved in cellular
75 pathways such as growth factor signaling, apoptosis, differentiation, and
76 cellular senescence^{13,14}. Thus, *miR34a* is a crucial factor in mediating
77 activated TP53 response and, the fact that it is often deleted or down-
78 regulated in human cancers indicates, its tumor suppressive effect and makes
79 it a valuable prognostic marker¹⁵⁻¹⁹. Reduced *miR34a* transcription is
80 mediated via epigenetic regulation in many solid tumors, including colorectal-,
81 pancreatic-, and ovarian cancer²⁰, as well as numerous types of
82 hematological malignancies²¹. In addition, *miR34a* has been shown to be
83 transcriptionally regulated via TP53 homologs, TP63 and TP73, other
84 transcription factors, e.g. STAT3 and MYC, and, in addition, post-
85 transcriptionally through miRNA sponging by the NEAT1 lncRNA²²⁻²⁶. Despite
86 these findings, the mechanisms underlying *miR34a* regulation in the context
87 of oncogenesis have not yet been fully elucidated.

88

89 Studies across multiple cancer types have reported a decrease in oncogenic
90 phenotypes when *miR34a* expression is induced in a *TP53*-null background,
91 although endogenous mechanisms for achieving this have not yet been

92 discovered^{18,27-30}. In addition, previous reports from large-scale studies
93 interrogating global TP53-mediated regulation of lncRNAs have identified a
94 lncRNA (known as RP3-510D11.2 and LINC01759) originating in the
95 antisense orientation from the *miR34a* locus that is induced upon numerous
96 forms of cellular stress³¹⁻³⁵. Despite this, none of these studies have
97 functionally characterized this transcript, which we name *Long-Non-coding*
98 *Transcriptional Activator of MiR34a* (lncTAM34a). In this study we functionally
99 characterize the *lncTAM34a* transcript, and find that it positively regulates
100 *miR34a* expression resulting in a decrease of several tumorigenic
101 phenotypes. Furthermore, we find that *lncTAM34a*-mediated up-regulation of
102 *miR34a* is sufficient to induce endogenous cellular mechanisms counteracting
103 several types of stress stimuli in a *TP53*-deficient background. Finally, similar
104 to the functional roles of antisense transcription at protein-coding genes, we
105 identify a rare example of an antisense RNA capable of regulating a cancer-
106 associated miRNA.

107

108 **Results**

109

110 ***lncTAM34a* is a broadly expressed non-coding transcript whose levels**
111 **correlate with *miR34a* expression**

112

113 *lncTAM34a* is transcribed in a “head-to-head” orientation with approximately
114 100 base pair overlap with the *miR34a* host gene (HG) (**Fig. 1a**). Due to the
115 fact that sense/antisense pairs can be both concordantly and discordantly
116 expressed, we sought to evaluate this relationship in the case of *miR34a* HG
117 and its asRNA. Using a diverse panel of cancer cell lines, we detected co-
118 expression of both the *miR34a* HG and *lncTAM34a* (**Fig. 1b**). We used cell
119 lines with a known *TP53* status in the panel due to previous reports

120 that *miR34a* and *lncTAM34a* are known downstream targets of TP53. These
121 results indicate that *miR34a* HG and *lncTAM34a* are co-expressed and that
122 their expression levels correlate with *TP53* status, with *TP53*^{-/-} cells tending to
123 have decreased or undetectable expression of both transcripts.

124

125 We next sought to analyze primary cancer samples to examine whether a
126 correlation between *lncTAM34a* and *miR34a* expression levels could be
127 identified. We utilized RNA sequencing data from The Cancer Genome Atlas
128 (TCGA) after stratifying patients by cancer type, *TP53* status, and, in the case
129 of breast cancer, cancer subtypes. The results indicate that *lncTAM34a*
130 and *miR34a* expression are strongly correlated in the vast majority of cancer
131 types examined, both in the presence and absence of wild-type *TP53* (**Fig.**
132 **1c, Supplementary Figure 1A**). The results also further confirm that the
133 expression levels of both *miR34a* and *lncTAM34a* are significantly reduced in
134 patients with nonsynonymous *TP53* mutations (**Supplementary Figure 1B**).

135

136 Next, we aimed to gain a thorough understanding of *lncTAM34a*'s molecular
137 characteristics and cellular localization. To experimentally determine the 3'
138 termination site for the *lncTAM34a* transcript we performed 3' rapid
139 amplification of cDNA ends (RACE) using the U2OS osteosarcoma cell line
140 that exhibited high endogenous levels of *lncTAM34a* in the cell panel
141 screening. Sequencing the cloned cDNA indicated that the transcripts 3'
142 transcription termination site is 525 bp upstream of the *lncTAM34a* transcript's
143 annotated termination site (**Fig. 1d**). Next, we characterized the *lncTAM34a* 5'
144 transcription start site by carrying out a primer walk assay, i.e. a common

145 reverse primer was placed in exon 2 and forward primers were gradually
146 staggered upstream of *lncTAM34a*'s annotated start site (**Supplementary**
147 **Figure 2A**). Our results indicated that the 5' start site for *lncTAM34a* is in fact
148 approximately 90 bp (F11 primer) to 220 bp (F12 primer) upstream of the
149 annotated start site (**Fig. 1e**). Polyadenylation status was evaluated via cDNA
150 synthesis with either random nanomers or oligo(DT) primers followed by semi-
151 quantitative PCR which showed that the *lncTAM34a* is polyadenylated
152 although the unspliced form seems to only be present in a polyadenylation
153 negative state (**Supplementary Figure 2B**). Furthermore, we investigated the
154 propensity of *lncTAM34a* to be alternatively spliced in U2OS cells, using PCR
155 cloning followed by sequencing and found that the transcript is post-
156 transcriptionally spliced to form multiple isoforms (**Supplementary Figure**
157 **2C**). In order to evaluate the subcellular localization of *lncTAM34a*, we made
158 use of RNA sequencing data from five cancer cell lines included in the
159 ENCODE³⁶ project that had been fractionated into cytosolic and nuclear
160 fractions. The analysis revealed that the *lncTAM34a* transcript primarily
161 localizes to the nucleus with only a minor fraction in the cytosol
162 (**Supplementary Figure 2D**).

163
164 Lastly, we utilized several approaches to evaluate the coding potential of
165 the *lncTAM34a* transcript. The Coding-Potential Assessment Tool is a
166 bioinformatics-based tool that uses a logistic regression model to evaluate
167 coding-potential by examining open reading frame (ORF) length, ORF
168 coverage, Fickett score, and hexamer score³⁷. Results indicated
169 that *lncTAM34a* has a similar low coding capacity to known non-coding

170 transcripts such as *HOTAIR* and *XIST* (Fig. 1F). We further confirmed these
171 results using the Coding-Potential Calculator that uses a support vector
172 machine-based classifier and accesses an alternate set of discriminatory
173 features (**Supplementary Figure 2E**)³⁸. Finally, we downloaded mass
174 spectrometry spectra for 11 cancer cell lines³⁹, 7 of which were also present
175 in the cell line panel above (**Fig. 1b**), and searched it against a database of
176 human protein sequences which also contained the 6 frame translation of
177 *IncTAM34a*. However, we did not manage to detect any peptides matching
178 the sequence in any of the 11 cell lines. Taken together our results indicate
179 that *IncTAM34a* is not a coding transcript and that it is not translated to any
180 significant degree.

181

182 **TP53-mediated regulation of *IncTAM34a* expression**

183 *miR34a* is a known downstream target of TP53 and has been previously
184 shown to exhibit increased expression within multiple contexts of cellular
185 stress. Several global analyses of TP53-regulated lncRNAs have also shown
186 *IncTAM34a* to be induced upon TP53 activation³¹⁻³⁵. To confirm these results
187 in our biological systems, we treated HEK293T, embryonic kidney cells, and
188 HCT116, colorectal cancer cells, with the DNA damaging agent doxorubicin to
189 activate TP53. QPCR-mediated measurements of both *miR34a* HG and
190 *IncTAM34a* indicated that their expression levels were increased in response
191 to doxorubicin treatment in both cell lines (**Fig. 2a**). To assess whether
192 TP53 was responsible for the increase in *IncTAM34a* expression upon DNA
193 damage, we treated *TP53^{+/+}* and *TP53^{-/-}* HCT116 cells with increasing
194 concentrations of doxorubicin and monitored the expression of

195 both *miR34a* HG and *lncTAM34a*. We observed a dose-dependent increase
196 in both *miR34a* HG and *lncTAM34a* expression levels with increasing
197 amounts of doxorubicin, revealing that these two transcripts are co-regulated,
198 although, this effect was largely abrogated in *TP53^{-/-}* cells (**Fig. 2b**). These
199 results indicate that TP53 activation increases *lncTAM34a* expression upon
200 DNA damage. Nevertheless, *TP53^{-/-}* cells also showed a dose-dependent
201 increase in both *miR34a* HG and *lncTAM34a*, suggesting that additional
202 factors, other than TP53 are capable of initiating an increase in expression of
203 both of these transcripts upon DNA damage.

204

205 The head-to-head orientation of *miR34a* HG and *lncTAM34a*, suggests that
206 transcription is initiated from a single promoter in a bi-directional manner (**Fig**
207 **1a**). To investigate whether *miR34a* HG and *lncTAM34a* are transcribed from
208 the same promoter as divergent transcripts, we cloned the previously reported
209 *miR34a* HG promoter, a ~300 bp region including the TP53 binding site and
210 the majority of the first exon of both transcripts, into a luciferase/renilla dual
211 reporter vector (**Supplementary Figure 3A-B**)¹². We hereafter refer to this
212 construct as p1. Upon transfection of p1 into HCT116 and HEK293T cell lines
213 we observed increases in both luciferase and renilla indicating
214 that *miR34a* HG and *lncTAM34a* expression can be regulated by a single
215 promoter contained within the p1 construct (**Fig. 2c**).

216

217 ***lncTAM34a* facilitates *miR34a* induction in response to DNA damage**
218 We hypothesized that *lncTAM34a* may regulate *miR34a* HG levels and, in
219 addition, that the overlapping regions of the sense and antisense transcripts

may mediate this regulation. Knockdown of endogenous *lncTAM34a* is complicated by its various isoforms (**Supplementary Figure 2C**). For this reason, we utilized the p1 construct to evaluate the regulatory role of *lncTAM34a* on *miR34a* HG. Accordingly, we first co-transfected the p1 construct, containing the overlapping region of the two transcripts, and two different short hairpin (sh) RNAs targeting renilla into HEK293T cells and subsequently measured luciferase and renilla expression. The results indicated that shRNA-mediated knock-down of the p1-renilla transcript (corresponding to *lncTAM34a*) caused p1-luciferase (corresponding to *miR34a* HG) levels to concomitantly decrease (**Supplementary Figure 3C**). The results suggest that *lncTAM34a* may positively regulate levels of *miR34a* HG and that the transcriptional product of *lncTAM34a* within the p1 construct contributes to inducing a *miR34a* response. To further support these conclusions and better understand the role of *lncTAM34a* during TP53 activation, *TP53^{+/+}* HCT116 cells were co-transfected with p1 and shRNA renilla (2.1) and subsequently treated with increasing doses of doxorubicin. Again, the results showed a concomitant reduction in luciferase levels upon knock-down of p1-renilla i.e. the *lncTAM34a* corresponding segment of the p1 transcript (**Fig. 2d**). Furthermore, the results showed that in the absence of p1-renilla the expected induction of p1-luciferase in response to TP53 activation by DNA damage is abrogated. Collectively these results indicate that *lncTAM34a* positively regulates *miR34a* expression and furthermore, suggests that it is crucial for an appropriate TP53-mediated *miR34a* response to DNA damage.

244

245 ***IncTAM34a* can regulate *miR34a* host gene independently of *TP53***

246 Despite the fact that TP53 regulates *miR34a* HG and *IncTAM34a* expression,
247 our results showed that other factors are also able to regulate this locus (**Fig.**
248 **2b**). Utilizing a lentiviral system, we stably over-expressed the *IncTAM34a*
249 transcript in three *TP53*-null cell lines, PC3 (prostate cancer), Saos2
250 (osteogenic sarcoma), and Skov3 (ovarian adenocarcinoma). We first
251 analyzed the levels of *IncTAM34a* in these stable cell lines, compared to
252 HEK293T cells, which have high endogenous levels of *IncTAM34a*. On
253 average, the over-expression was approximately 30-fold higher in the over-
254 expression cell lines than in HEK293T cells, roughly corresponding to
255 physiologically relevant levels in cells encountering a stress stimulus, such as
256 DNA damage (**Supplementary Figure 4A**). Analysis of *miR34a* levels in
257 the *IncTAM34a* over-expressing cell lines showed that this over-expression
258 resulted in a concomitant increase in the expression of *miR34a* in all three cell
259 lines (**Fig. 3a**). These results indicate that, in the absence of
260 *TP53*, *miR34a* expression may be rescued by activating *IncTAM34a*
261 expression.

262

263 *miR34a* has been previously shown to regulate cell cycle progression, with
264 *miR34a* induction causing G1 arrest^{12,40}. Cell cycle analysis via determination
265 of DNA content showed a significant increase in G1 phase cells and a
266 concomitant decrease in G2 phase cells in the PC3 and Skov3 *IncTAM34a*
267 over-expressing cell lines, indicating G1 arrest (**Fig. 3b**). The effects of
268 *miR34a* on the cell cycle are mediated by its ability to target cell cycle
269 regulators such as cyclin D1 (*CCND1*)⁴¹. Quantification of both *CCND1* RNA

270 expression (**Supplementary Figure 4B**) and protein levels (**Supplementary**
271 **Figure 4C**) in the PC3 *lncTAM34a* over-expressing cell line showed a
272 significant decrease of *CCND1* levels compared to the mock control.
273 Collectively, these results indicate that *lncTAM34a*-mediated induction of
274 *miR34a* is sufficient to result in the corresponding *miR34a*-directed effects on
275 cell cycle.

276

277 *miR34a* is also a well-known inhibitor of cellular growth via its ability to
278 negatively regulate growth factor signaling. Furthermore, starvation has been
279 shown to induce *miR34a* expression causing inactivation of numerous pro-
280 survival growth factors¹³. We further interrogated the effects of *lncTAM34a*
281 over-expression by monitoring the growth of the PC3 stable cell lines in both
282 normal and starvation conditions via confluency measurements over a 35-
283 hour period. Under normal growth conditions there is a small but significant
284 reduction ($P = 3.0\text{e-}8$; linear regression, **Fig. 3c**) in confluency in the
285 *lncTAM34a* over-expressing cell lines compared to mock control. However,
286 these effects on cell growth are drastically increased in starvation conditions
287 ($P = 9.5\text{e-}67$; linear regression; **Fig. 3c**). This is in agreement with our
288 previous results, and suggests that *lncTAM34a*-mediated increases
289 in *miR34a* expression are crucial under conditions of stress and necessary for
290 the initiation of an appropriate cellular response. In summary, we find that
291 over-expression of *lncTAM34a* is sufficient to increase *miR34a* expression
292 and gives rise to known phenotypes observed upon induction of *miR34a*.

293

294 ***lncTAM34a* transcriptionally activates *miR34a* host gene**

295 Antisense RNAs have been reported to mediate their effects both via
296 transcriptional and post-transcriptional mechanisms. Due to the fact that
297 *miR34a* expression is undetected in wild type PC3 cells (**Fig. 1b**) but, upon
298 over-expression of *lncTAM34a*, increases to detectable levels, we
299 hypothesized that *lncTAM34a* is capable of regulating *miR34a* expression via
300 a transcriptional mechanism. To ascertain if this is actually the case, we
301 performed chromatin immunoprecipitation (ChIP) for phosphorylated
302 polymerase II (polII) at the *miR34a* HG promoter in both *lncTAM34a* over-
303 expressing and mock control cell lines. Our results indicated a clear increase
304 in phosphorylated polII binding at the *miR34a* promoter upon *lncTAM34a*
305 over-expression indicating the ability of *lncTAM34a* to transcriptionally
306 regulate *miR34a* levels (**Fig. 3d**).
307

308 **Low *lncTAM34a* expression levels are associated with decreased
309 survival**

310 As *TP53* mutations and low expression of *miR34a* have been associated with
311 worse prognosis in cancer, we compared survival rates of samples with low
312 expression of *lncTAM34a* (bottom 10th percentile) to control samples in 17
313 cancer types from TCGA (**Supplementary Figure 5**)¹⁷⁻¹⁹. To correct for the
314 effect of *TP53* mutations we focused on non-*TP53* mutated samples, and
315 noted a worse survival for the low expression group in several cancers. This
316 effect was most pronounced in papillary kidney cancer (unadjusted
317 $P=0.00095$; **Fig. 4a**). By systematically comparing 5-year survival probabilities
318 between the low expression group and the control group for each cancer we
319 found a median reduction of 5-year survival probability of 9.6% ($P=0.083$;
320 Wilcoxon signed rank test; **Fig. 4b**). Furthermore, we found that *lncTAM34a*

322 expression showed similar patterns in terms of direction and strength of
323 association with 5-year survival probability as *miR34a* expression ($r=0.57$,
324 $P=0.037$) and *TP53* mutations ($r=0.80$, $P=0.00054$) across the different
325 cancer types (**Fig. 4b**). Although these results do not implicate any causal
326 relationship, they do indicate a striking similarity between the association of
327 worse prognosis and *TP53* mutations, low *miR34a*, and low *lncTAM34a*
328 expression.

329

330 **Discussion**

331
332 Multiple studies have previously shown asRNAs to be crucial for the
333 appropriate regulation of cancer-associated protein-coding genes and that
334 their dysregulation can lead to perturbation of tumor suppressive and
335 oncogenic pathways, as well as, cancer-related phenotypes^{6,7,42,43}. Here we
336 show that asRNAs are also capable of regulating cancer-associated miRNAs
337 resulting in similar consequences as protein-coding gene dysregulation (**Fig.**
338 **4**). Interestingly, we show that, both in the presence and absence of
339 *TP53*, *lncTAM34a* provides an additional regulatory level to control *miR34a*
340 expression in both homeostasis and upon encountering various forms of
341 cellular stress. Furthermore, we find that *lncTAM34a*-mediated increase in
342 *miR34a* expression is sufficient to drive the appropriate cellular responses to
343 these stress stimuli (**Fig. 2d and Fig. 3c**). Previous studies have exploited
344 various molecular biology methods to up-regulate *miR34a* expression in cells
345 lacking wild type *TP53*^{18,27-30}. In this study, we demonstrate a novel,
346 endogenous mechanism of *miR34a* regulation that has similar phenotypic
347 outcomes as has been previously shown for *miR34a* induction in a *TP53*
348 deficient background.

349

350 In agreement with previous studies, we demonstrate that upon encountering
351 various types of cellular stress, TP53 in concert with additional factors initiates
352 transcription at the *miR34a* locus, thus increasing the levels of *lncTAM34a*
353 and *miR34a*³¹⁻³⁵. We found that over-expression of *lncTAM34a* leads to
354 recruitment of polII to the *miR34a* promoter and hypothesize that *lncTAM34a*
355 may provide positive feedback for *miR34a* expression whereby it serves as a

356 scaffold for the recruitment of additional factors that facilitate polII-mediated
357 transcription. In this manner, *miR34a* expression is induced, driving a shift
358 towards senescence, a reduction in growth factor signaling, and in some
359 cases, apoptosis. On the other hand, in cells without functional TP53, other
360 factors, which typically act independently or in concert with TP53, may initiate
361 transcription of the *miR34a* locus. Due to the fact that *lncTAM34a* can alter
362 *miR34a* expression in these cells, we suggest that it is interacting with one of
363 these additional factors, possibly recruiting it to the *miR34a* locus in order to
364 drive *miR34a* transcription, similar to mechanisms described for other
365 lncRNAs⁴⁴⁻⁴⁶. The head-to-head orientation of the *miR34a* HG and
366 *lncTAM34a* causes sequence complementarity between the RNA and the
367 promoter DNA, making targeting by direct binding an attractive mechanism.
368 Previous reports have also illustrated the ability of asRNAs to form hybrid
369 DNA:RNA R-loops and, thus, facilitate an open chromatin structure and the
370 transcription of the sense gene⁴⁷. The fact that the p1 construct only contains
371 a small portion (~300 bp) of the *lncTAM34a* transcript indicates that this
372 portion is sufficient to give rise to at least a partial *miR34a* inducing response
373 and therefore, that *lncTAM34a* may be able to facilitate *miR34a* expression
374 independent of additional factors (**Fig 2d, Supplementary Figure 3C**).
375 Nevertheless, further work will need to be performed to explore the
376 mechanism whereby *lncTAM34a* regulates *miR34a* gene expression.

377

378 An antisense transcript arising from the *miR34a* locus, *Lnc34a*, has been
379 previously reported to negatively regulate the expression of *miR34a*⁴⁸.
380 Although the *Lnc34a* and *lncTAM34a* transcripts share some sequence

381 similarity, we believe them to be separate RNAs that are, potentially, different
382 isoforms of the same gene. We utilized CAGE and RNAseq data from the
383 ENCODE project to evaluate the presence of *IncTAM34a* and *Lnc34a* in 28
384 and 36 commonly used cancer cell lines, respectively. Although the results
385 show the presence of *IncTAM34a* in these cell lines, we find no evidence for
386 *Lnc34a* transcription (**Supplementary Figure 6-7**). These results are in line
387 with the findings of Wang et al. indicating that *Lnc34a* is highly expressed in
388 colon cancer stem cell spheres compared to all other cell types used in their
389 study and may not be broadly expressed in other tissues or tumor types. The
390 fact that *IncTAM34a* and *Lnc34a* would appear to have opposing roles in their
391 regulation of *miR34a*, further underlines the complexity of the regulation at
392 this locus.

393

394 Clinical trials utilizing *miR34a* replacement therapy have previously been
395 conducted but, disappointingly, were terminated after adverse side effects of
396 an immunological nature were observed in several of the patients¹⁴. Although
397 it is not presently clear if these side effects were caused by *miR34a* or the
398 liposomal carrier used to deliver the miRNA, the multitude of evidence
399 indicating *miR34a*'s crucial role in oncogenesis still makes its therapeutic
400 induction an interesting strategy and needs further investigation. Our results
401 indicate an association between survival probability and low *IncTAM34a*
402 expression making it an attractive candidate for controlled preclinical studies.
403 Due to the *IncTAM34a*-mediated positive feedback on *miR34a* expression,
404 initiation of this feedback mechanism may provide a sustained *miR34a*
405 induction in a relatively more robust manner than *miR34a* replacement alone.

406 In summary, our results have identified *lncTAM34a* as a vital component in
407 the regulation of *miR34a* and its particular importance in typical examples of
408 cellular stress encountered in cancer. On a broader level, the conclusions
409 drawn in this study provide an example of asRNA-mediated regulation of a
410 clinically relevant cancer-associated miRNA and contribute to fundamental
411 knowledge concerning *miR34a* regulation.

412

413 **Materials and Methods**

414 **Cell Culture**

415 All cell lines were cultured at 5% CO₂ and 37°C with HEK293T, Saos2, and
416 Skov3 cells cultured in DMEM high glucose (GE Healthcare Life Sciences,
417 Hyclone, Amersham. UK, Cat# SH30081), HCT116 and U2OS cells in
418 McCoy's 5a (ThermoFisher Scientific, Pittsburgh, MA, USA. Cat# SH30200),
419 and PC3 cells in RPMI (GE Healthcare Life Sciences, Hyclone, Cat#
420 SH3009602) and 2 mM L-glutamine (GE Healthcare Life Sciences, Hyclone,
421 Cat# SH3003402). All growth mediums were supplemented with 10% heat-
422 inactivated FBS (ThermoFisher Scientific, Gibco, Cat# 12657029) and 50
423 µg/ml of streptomycin (ThermoFisher Scientific, Gibco, Cat# 15140122) and
424 50 µg/ml of penicillin (ThermoFisher Scientific, Gibco, Cat# 15140122). All cell
425 lines were purchased from ATCC, tested negative for mycoplasma, and their
426 identity was verified via STR profiling.

427

428 **Bioinformatics, Data Availability, and Statistical Testing**

429 The USCS genome browser⁴⁹ was utilized for the bioinformatic evaluation of
430 antisense transcription utilizing the RefSeq⁵⁰ gene annotation track.

431
432 All raw experimental data, code used for analysis, and supplementary
433 methods are available for review at⁵¹ and are provided as an R package. All
434 analysis took place using the R statistical programming language⁵² using
435 external packages that are documented in the package associated with this
436 article⁵³⁻⁶⁴. The package facilitates replication of the operating system and
437 package versions used for the original analysis, reproduction of each
438 individual figure and figure supplement included in the article, and easy review
439 of the code used for all steps of the analysis, from raw-data to figure.

440
441 The significance threshold (alpha) in this study was set to 0.05. Statistical
442 testing was performed using an unpaired two sample Student's two-sided t-
443 test unless otherwise specified. Data was either approximated to be normally
444 distributed or transformed to be so in cases where a parametric test was
445 utilized. In addition, variance was not assumed to be equal between groups
446 and, therefore, the Welch (or Satterthwaite) approximation to the degrees of
447 freedom was used.

448
449 **Coding Potential**
450 Protein-coding capacity was evaluated using the Coding-potential
451 Assessment Tool³⁷ and Coding-potential Calculator³⁸ with default settings.
452 Transcript sequences for use with Coding-potential Assessment Tool were
453 downloaded from the UCSC genome browser using the Ensembl
454 accessions: *HOTAIR* (ENST00000455246), *XIST* (ENST00000429829), β-
455 actin (ENST00000331789), Tubulin (ENST00000427480),

456 and *MYC* (ENST00000377970). Transcript sequences for use with Coding-
457 potential Calculator were downloaded from the UCSC genome browser using
458 the following IDs: *HOTAIR* (uc031qho.1), β-actin (uc003soq.4).

459

460 **Peptide identification in MS/MS spectra**

461 Orbitrap raw MS/MS files for 11 human cell lines were downloaded from the
462 PRIDE repository (PXD002395; ³⁹) converted to mzML format using
463 msConvert from the ProteoWizard tool suite⁶⁵. Spectra were then searched
464 using MSGF+ (v10072)⁶⁶ and Percolator (v2.08)⁶⁷. All searches were done
465 against the human protein subset of Ensembl 75 in the Galaxy platform⁶⁸
466 supplemented with the 6 frame translation of both the annotated
467 (LOC102724571; hg38) and PCR cloned sequence of *IncTAM34a*
468 (supplementary data; ⁵¹). MSGF+ settings included precursor mass tolerance
469 of 10 ppm, fully-tryptic peptides, maximum peptide length of 50 amino acids
470 and a maximum charge of 6. Fixed modification was carbamidomethylation on
471 cysteine residues; a variable modification was used for oxidation on
472 methionine residues. Peptide Spectral Matches found at 1% FDR (false
473 discovery rate) were used to infer peptide identities. The output from all
474 searches are available in⁵¹.

475

476 **shRNAs**

477 shRNA-expressing constructs were cloned into the U6M2 construct using the
478 BgIII and KpnI restriction sites as previously described⁶⁹. shRNA constructs
479 were transfected using Lipofectamine 2000 or 3000 (ThermoFisher Scientific,
480 Cat# 12566014 and L3000015). The sequences targeting renilla is as follows:

481 shRenilla 1.1 (AAT ACA CCG CGC TAC TGG C), shRenilla 2.1 (TAA CGG
482 GAT TTC ACG AGG C).

483

484 **Bi-directional Promoter Cloning**

485 The overlapping region (p1) corresponds with the sequence previously
486 published as the TP53 binding site in¹² which we synthesized, cloned into the
487 pLucRluc construct⁷⁰, and sequenced to verify its identity.

488

489 **Promoter Activity**

490 Cells were co-transfected with the p1 renilla/firefly bidirectional promoter
491 construct⁷⁰ and GFP by using Lipofectamine 2000 (Life Technologies, Cat#
492 12566014). The expression of GFP and luminescence was measured 24
493 hours post transfection by using the Dual-Glo Luciferase Assay System
494 (Promega, Cat# E2920) and detected by the GloMax-Multi+ Detection System
495 (Promega, Cat# SA3030). The expression of luminescence was normalized to
496 GFP.

497

498 **Generation of U6-expressed *IncTAM34a* Lentiviral Constructs**

499 The U6 promoter was amplified from the U6M2 cloning plasmid⁶⁹ and ligated
500 into the Not1 restriction site of the pHIV7-IMPDH2 vector⁷¹. *IncTAM34a* was
501 PCR amplified and subsequently cloned into the Nhe1 and Pac1 restriction
502 sites in the pHIV7-IMPDH2-U6 plasmid.

503

504 **Lentiviral Particle production, infection, and selection**

505 Lentivirus production was performed as previously described in⁷¹. Briefly,

506 HEK293T cells were transfected with viral and expression constructs using
507 Lipofectamine 2000 (ThermoFisher Scientific, Cat# 12566014), after which
508 viral supernatants were harvested 48 and 72 hours post-transfection. Viral
509 particles were concentrated using PEG-IT solution (Systems Biosciences,
510 Palo Alto, CA, USA. Cat# LV825A-1) according to the manufacturer's
511 recommendations. HEK293T cells were used for virus titration and GFP
512 expression was evaluated 72hrs post-infection via flow cytometry (LSRII, BD
513 Biosciences, San Jose, CA, USA) after which TU/ml was calculated.

514

515 Stable lines were generated by infecting cells with a multiplicity of infection of
516 1 and subsequently initiating 1-2 µM mycophenolic acid-based (Merck,
517 Kenilworth, NJ, USA. Cat# M5255) selection 48-72 hours post-infection. Cells
518 were expanded as the selection process was monitored via flow cytometry
519 analysis (LSRII, BD Biosciences) of GFP and selection was terminated once
520 > 90% of the cells were GFP positive. Quantification of *lncTAM34a* over-
521 expression and *miR34a* was performed in biological quintuplet for all cell
522 lines.

523

524 **Western Blotting**

525 Samples were lysed in 50 mM Tris-HCl (Sigma Aldrich, St. Louis, MO, USA.
526 Cat# T2663), pH 7.4, 1% NP-40 (Sigma Aldrich, Cat# I8896), 150 mM NaCl
527 (Sigma Aldrich, Cat# S5886), 1 mM EDTA (Promega, Madison, WI, USA.
528 Cat# V4231), 1% glycerol (Sigma Aldrich, Cat# G5516), 100 µM vanadate
529 (Sigma Aldrich, Cat# S6508), protease inhibitor cocktail (Roche Diagnostics,
530 Basel, Switzerland, Cat# 004693159001) and PhosSTOP (Roche

531 Diagnostics, Cat# 04906837001). Lysates were subjected to SDS-PAGE and
532 transferred to PVDF membranes. The proteins were detected by western blot
533 analysis by using an enhanced chemiluminescence system (Western
534 Lightning-ECL, PerkinElmer, Waltham, MA, USA. Cat# NEL103001EA).
535 Antibodies used were specific for CCND1 1:1000 (Cell Signaling, Danvers,
536 MA, USA. Cat# 2926), and GAPDH 1:5000 (Abcam, Cambridge, UK, Cat#
537 ab9485). All western blot quantifications were performed using ImageJ⁷².
538

539 **RNA Extraction and cDNA Synthesis**

540 For downstream SYBR green applications, RNA was extracted using the
541 RNeasy mini kit (Qiagen, Venlo, Netherlands, Cat# 74106) and subsequently
542 treated with DNase (Ambion Turbo DNA-free, ThermoFisher Scientific, Cat#
543 AM1907). 500ng RNA was used for cDNA synthesis using MuMLV
544 (ThermoFisher Scientific, Cat# 28025013) and a 1:1 mix of oligo(dT) and
545 random nanomers.

546

547 For analysis of miRNA expression with Taqman, samples were isolated with
548 TRIzol reagent (ThermoFisher Scientific, Cat# 15596018) and further
549 processed with the miRNeasy kit (Qiagen, Cat# 74106). cDNA synthesis was
550 performed using the TaqMan MicroRNA Reverse Transcription Kit
551 (ThermoFisher Scientific, Cat# 4366597) using the corresponding oligos
552 according to the manufacturer's recommendations.

553

554 **QPCR and PCR**

555 PCR was performed using the KAPA2G Fast HotStart ReadyMix PCR Kit

556 (Kapa Biosystems, Wilmington, MA, USA, Cat# KK5601) with corresponding
557 primers. QPCR was carried out using KAPA 2G SYBRGreen (Kapa
558 Biosystems, Cat# KK4602) using the Applied Biosystems 7900HT machine
559 with the cycling conditions: 95 °C for 3 min, 95 °C for 3 s, 60 °C for 30 s.

560

561 QPCR for miRNA expression analysis was performed according to the primer
562 probe set manufacturers recommendations (ThermoFisher Scientific) and
563 using the TaqMan Universal PCR Master Mix (ThermoFisher Scientific, Cat#
564 4304437) with the same cycling scheme as above. Primer and probe sets for
565 TaqMan were also purchased from ThermoFisher Scientific (Life
566 Technologies at time of purchase, TaqMan® MicroRNA Assay, hsa-miR-34a,
567 human, Cat# 4440887, Assay ID: 000426 and Control miRNA Assay, RNU48,
568 human, Cat# 4440887, Assay ID: 001006).

569

570 The $\Delta\Delta Ct$ method was used to quantify gene expression. All QPCR-based
571 experiments were performed in at least technical duplicate. Primers for all
572 PCR-based experiments are listed in **Supplementary Document 2** and
573 arranged by figure.

574

575 **Cell Cycle Distribution**

576 Cells were washed in PBS and fixed in 4% paraformaldehyde at room
577 temperature overnight. Paraformaldehyde was removed, and cells were re-
578 suspended in 95% EtOH. The samples were then rehydrated in distilled
579 water, stained with DAPI and analyzed by flow cytometry on a LSRII (BD
580 Biosciences) machine. Resulting cell cycle phases were quantified using the

581 ModFit software (Verity Software House, Topsham, ME, USA). Experiments
582 were performed in biological quadruplet (PC3) or triplicate (Skov3). The log₂
583 fraction of cell cycle phase was calculated for each replicate and a two
584 sample t-test was utilized for statistical testing.

585

586 **3' Rapid Amplification of cDNA Ends**

587 3'-RACE was performed as described as previously in⁸. Briefly, U2OS cell
588 RNA was polyA-tailed using yeast polyA polymerase (ThermoFisher
589 Scientific, Cat# 74225Z25KU) after which cDNA was synthesized using
590 oligo(dT) primers. Nested-PCR was performed first using a forward primer in
591 *IncTAM34a* exon 1 and a tailed oligo(dT) primer followed by a second PCR
592 using an alternate *IncTAM34a* exon 1 primer and a reverse primer binding to
593 the tail of the previously used oligo(dT) primer. PCR products were gel
594 purified and cloned the Strata Clone Kit (Agilent Technologies, Santa Clara,
595 CA, USA. Cat# 240205), and sequenced.

596

597 **Chromatin Immunoprecipitation**

598 The ChIP was performed as previously described in⁸ with the following
599 modifications. Cells were crosslinked in 1% formaldehyde (Merck, Cat#
600 1040039025), quenched with 0.125M glycine (Sigma Aldrich, Cat# G7126),
601 and lysed in cell lysis buffer comprised of: 5mM PIPES (Sigma Aldrich, Cat#
602 80635), 85mM KCL (Merck, Cat# 4936), 0.5% NP40 (Sigma Aldrich, Cat#
603 I8896), protease inhibitor (Roche Diagnostics, Cat# 004693159001). Samples
604 were then sonicated in 50mM TRIS-HCL pH 8.0 (Sigma Aldrich, MO, USA,
605 Cat# T2663) 10mM EDTA (Promega, WI, USA, Cat# V4231), 1% SDS

606 (ThermoFisher Scientific, Cat# AM9822), and protease inhibitor (Roche
607 Diagnostics, Cat# 004693159001) using a Bioruptor Sonicator (Diagenode,
608 Denville, NJ, USA). Samples were incubated over night at 4°C with
609 the polII antibody (Abcam, Cat# ab5095) and subsequently pulled down with
610 Salmon Sperm DNA/Protein A Agarose (Millipore, Cat# 16-157) beads. DNA
611 was eluted in an elution buffer of 1% SDS (ThermoFisher Scientific, Cat#
612 AM9822) 100mM NaHCO3 (Sigma Aldrich, Cat# 71631), followed by reverse
613 crosslinking, RNaseA (ThermoFisher Scientific, Cat# 1692412) and protease
614 K (New England Biolabs, Ipswich, MA, USA, Cat# P8107S) treatment. The
615 DNA was eluted using Qiagen PCR purification kit (Cat# 28106) and
616 quantified via QPCR. QPCR was performed in technical duplicate using the
617 standard curve method and reported absolute values. The fraction of input
618 was subsequently calculated using the mean of the technical replicates
619 followed by calculating the fold over the control condition. Statistical testing
620 was performed using 4 biological replicates with the null hypothesis that the
621 true log2 fold change values were equal to zero.

622

623 **Confluency Analysis**

624 Cells were incubated in the Spark Multimode Microplate (Tecan, Männedorf,
625 Switzerland) reader for 48 hours at 37°C with 5% CO₂ in a humidity chamber
626 in either normal medium or HBSS (ThermoFisher Scientific, Cat# 14025092).
627 Confluency was measured every hour using bright-field microscopy and the
628 percentage of confluency was reported via the plate reader's inbuilt algorithm.
629 Percentage of confluency was normalized to the control sample in each
630 condition (shown in figure) and then ranked to move the data to a linear scale.

631 Using the mean of the technical duplicates in three biological replicates, the
632 rank was then used to construct a linear model, of the dependency of the rank
633 on the time and cell lines variables for each growth condition. Reported P
634 values are derived from the t-test, testing the null hypothesis that the
635 coefficient estimate of the cell line variable is equal to 0.

636

637 **Pharmacological Compounds**

638 Doxorubicin was purchased from Teva (Petah Tikva, Israel, cat. nr. 021361).

639

640 **Cellular Localization Analysis**

641 Quantified RNAseq data from 11 cell lines from the GRCh38 assembly was
642 downloaded from the ENCODE project database and quantifications for
643 *IncTAM34a* (ENSG00000234546), GAPDH (ENSG00000111640), and
644 MALAT1 (ENSG00000251562) were extracted. Cell lines for which data was
645 downloaded include: A549, GM12878, HeLa-S3, HepG2, HT1080, K562
646 MCF-7, NCI-H460, SK-MEL-5, SK-N-DZ, SK-N-SH. Initial exploratory analysis
647 revealed that several cell lines should be removed from the analysis due to a)
648 a larger proportion of GAPDH in the nucleus than cytoplasm or b) variation of
649 *IncTAM34a* expression is too large to draw conclusions, or c) they have no or
650 low (<6 TPM) *IncTAM34a* expression. Furthermore, only polyadenylated
651 libraries were used in the final analysis, due to the fact that the cellular
652 compartment enrichment was improved in these samples. All analyzed genes
653 are reported to be polyadenylated. In addition, only samples with 2 biological
654 replicates were retained. For each cell type, gene, and biological replicate the
655 fraction of transcripts per million (TPM) in each cellular compartment was

656 calculated as the fraction of TPM in the specific compartment by the total
657 TPM. The mean and standard deviation for the fraction was subsequently
658 calculated for each cell type and cellular compartment and this information
659 was represented in the final figure.

660

661 **CAGE Analysis**

662 All available CAGE data from the ENCODE project³⁶ for 36 cell lines was
663 downloaded from the UCSC genome browser⁴⁹ for genome version hg19. Of
664 these, 28 cell lines had CAGE transcription start sites (TSS) mapping to the
665 plus strand of chromosome 1 and in regions corresponding to 200 base pairs
666 upstream of the *Lnc34a* start site (9241796 - 200) and 200 base pairs
667 upstream of the GENCODE annotated *IncTAM34a* start site (9242263 + 200).
668 These cell lines included: HFDPC, H1-hESC, HMEpC, HAoEC, HPIEpC,
669 HSaVEC, GM12878, hMSC-BM, HUVEC, AG04450, hMSC-UC, IMR90,
670 NHDF, SK-N-SH_RA, BJ, HOB, HPC-PL, HAoAF, NHEK, HVMF, HWP, MCF-
671 7, HepG2, hMSC-AT, NHEM.f_M2, SkMC, NHEM_M2, and HCH. In total 74
672 samples were included. 17 samples were polyA-, 47 samples were polyA+,
673 and 10 samples were total RNA. In addition, 34 samples were whole cell, 15
674 enriched for the cytosolic fraction, 15 enriched for the nucleolus, and 15
675 enriched for the nucleus. All CAGE transcription start sites were plotted and
676 the RPKM of the individual reads was used to color each read to indicate their
677 relative abundance. In cases where CAGE TSS spanned identical regions,
678 the RPMKs of the regions were summed and represented as one CAGE TSS
679 in the figure. In addition, a density plot shows the distribution of the CAGE
680 reads in the specified interval.

681

682 **Splice Junction Analysis**

683 All available whole cell (i.e. non-fractionated) spliced read data originating
684 from the Cold Spring Harbor Lab in the ENCODE project³⁶ for 38 cell lines
685 was downloaded from the UCSC genome browser⁴⁹. Of these cell lines, 36
686 had spliced reads mapping to the plus strand of chromosome 1 and in the
687 region between the *Lnc34a* start (9241796) and transcription termination
688 (9257102) site (note that *IncTAM34a* resides totally within this region). Splice
689 junctions from the following cell lines were included in the final figure: A549,
690 Ag04450, Bj, CD20, CD34 mobilized, Gm12878, H1hesc, Haoaf, Haoec, Hch,
691 Helas3, Hepg2, Hfdpc, Hmec, Hmepc, Hmscat, Hmscbm, Hmscuc, Hob,
692 Hpcpl, Hpiepc, Hsavec, Hsomm, Huvec, Hvmf, Hwp, Imr90, Mcf7, Monocd14,
693 Nhdf, Nhek, Nhemfm2, Nhemm2, Nhlf, Skmc, and Sknsh. All splice junctions
694 were included in the figure and colored according to the number of reads
695 corresponding to each. In cases where identical reads were detected multiple
696 times, the read count was summed and represented as one read in the figure.

697

698 **TCGA Data Analysis**

699 RNAseq data and copy number data were downloaded from TCGA and
700 processed as described previously³⁴. Briefly, RNAseq data were aligned to
701 the human hg19 assembly and quantified using GENCODE (v19) annotated
702 HTSeq-counts and FPKM normalizations. Expression data from *miR34a* and
703 *IncTAM34a* (identified as RP3-510D11.2) were used for further analysis. Copy
704 number amplitudes for GENCODE genes were determined from segmented

705 copy-number data. Samples that were diploid for *IncTAM34a* were identified
706 as those samples that had copy number amplitudes between -0.1 and 0.1.

707

708 Somatic mutation data were downloaded from the Genomics Data Commons
709 data portal (GDC) as mutation annotation format (maf) files, called using
710 Mutect2 on 30/10/2017 (v7)⁷³.

711

712 Survival analysis was performed on TCGA vital state and follow-up data,
713 downloaded from GDC on 27/10/2017 using the R survival package⁶⁴.

714

715 **Acknowledgments**

716 The authors would like to kindly thank Martin Enge for his critical review of the
717 manuscript and fruitful discussions. Dan Grandér, who played a significant
718 role in the conceptualization and supervision of this project, sadly passed
719 away before the initial submission of the manuscript. May he rest in peace.

720

721 **Competing Interests**

722 The authors declare no competing interests.
723

724

725

726

727 **Funding**

728

729 This work has been supported by the Swedish Research Council [521-2012-
730 2037], Swedish Cancer Society [150768], Cancer Research Foundations of
731 Radiumhemmet [144063] and the Swedish Childhood Cancer Foundation
732 [PR2015-0009].

733

734 **Supplementary Information**

735 Supplementary information is available online at Cell Death and Disease's
736 website.

737

738

739 **References**

740

- 741 1 International Human Genome Sequencing, C. Finishing the euchromatic
742 sequence of the human genome. *Nature* **431**, 931-945,
743 doi:10.1038/nature03001 (2004).
- 744 2 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**,
745 101-108, doi:10.1038/nature11233 (2012).
- 746 3 Ozsolak, F. *et al.* Comprehensive polyadenylation site maps in yeast and
747 human reveal pervasive alternative polyadenylation. *Cell* **143**, 1018-1029,
748 doi:10.1016/j.cell.2010.11.020 (2010).
- 749 4 Chen, J. *et al.* Over 20% of human transcripts might form sense-antisense
750 pairs. *Nucleic acids research* **32**, 4812-4820, doi:10.1093/nar/gkh818
751 (2004).
- 752 5 Katayama, S. *et al.* Antisense transcription in the mammalian
753 transcriptome. *Science* **309**, 1564-1566, doi:10.1126/science.1112009
754 (2005).
- 755 6 Balbin, O. A. *et al.* The landscape of antisense gene expression in human
756 cancers. *Genome research* **25**, 1068-1079, doi:10.1101/gr.180596.114
757 (2015).
- 758 7 Yap, K. L. *et al.* Molecular interplay of the noncoding RNA ANRIL and
759 methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional
760 silencing of INK4a. *Molecular cell* **38**, 662-674,
761 doi:10.1016/j.molcel.2010.03.021 (2010).
- 762 8 Johnsson, P. *et al.* A pseudogene long-noncoding-RNA network regulates
763 PTEN transcription and translation in human cells. *Nature structural &*
- 764 *molecular biology* **20**, 440-446, doi:10.1038/nsmb.2516 (2013).
- 765 9 Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin
766 domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323,
767 doi:10.1016/j.cell.2007.05.022 (2007).
- 768 10 Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with
769 regulatory potency. *Nature* **495**, 333-338, doi:10.1038/nature11928
770 (2013).
- 771 11 Conley, A. B. & Jordan, I. K. Epigenetic regulation of human cis-natural
772 antisense transcripts. *Nucleic acids research* **40**, 1438-1445,
773 doi:10.1093/nar/gkr1010 (2012).
- 774 12 Raver-Shapira, N. *et al.* Transcriptional activation of miR-34a contributes
775 to p53-mediated apoptosis. *Molecular cell* **26**, 731-743,
776 doi:10.1016/j.molcel.2007.05.017 (2007).
- 777 13 Lal, A. *et al.* Capture of microRNA-bound mRNAs identifies the tumor
778 suppressor miR-34a as a regulator of growth factor signaling. *PLoS*
779 *genetics* **7**, e1002363, doi:10.1371/journal.pgen.1002363 (2011).
- 780 14 Slabakova, E., Culig, Z., Remsik, J. & Soucek, K. Alternative mechanisms of
781 miR-34a regulation in cancer. *Cell Death Dis* **8**, e3100,
782 doi:10.1038/cddis.2017.495 (2017).
- 783 15 Cheng, J. *et al.* The impact of miR-34a on protein output in hepatocellular
784 carcinoma HepG2 cells. *Proteomics* **10**, 1557-1572,
785 doi:10.1002/pmic.200900646 (2010).

- 786 16 Cole, K. A. *et al.* A functional screen identifies miR-34a as a candidate
787 neuroblastoma tumor suppressor gene. *Molecular cancer research : MCR*
788 **6**, 735-742, doi:10.1158/1541-7786.MCR-07-2102 (2008).
- 789 17 Gallardo, E. *et al.* miR-34a as a prognostic marker of relapse in surgically
790 resected non-small-cell lung cancer. *Carcinogenesis* **30**, 1903-1909,
791 doi:10.1093/carcin/bgp219 (2009).
- 792 18 Liu, C. *et al.* The microRNA miR-34a inhibits prostate cancer stem cells
793 and metastasis by directly repressing CD44. *Nature medicine* **17**, 211-215,
794 doi:10.1038/nm.2284 (2011).
- 795 19 Zenz, T. *et al.* miR-34a as part of the resistance network in chronic
796 lymphocytic leukemia. *Blood* **113**, 3801-3808, doi:10.1182/blood-2008-
797 08-172254 (2009).
- 798 20 Vogt, M. *et al.* Frequent concomitant inactivation of miR-34a and miR-
799 34b/c by CpG methylation in colorectal, pancreatic, mammary, ovarian,
800 urothelial, and renal cell carcinomas and soft tissue sarcomas. *Virchows
801 Archiv : an international journal of pathology* **458**, 313-322,
802 doi:10.1007/s00428-010-1030-5 (2011).
- 803 21 Chim, C. S. *et al.* Epigenetic inactivation of the miR-34a in hematological
804 malignancies. *Carcinogenesis* **31**, 745-750, doi:10.1093/carcin/bgq033
805 (2010).
- 806 22 Agostini, M. *et al.* Neuronal differentiation by TApx73 is mediated by
807 microRNA-34a regulation of synaptic protein targets. *Proceedings of the
808 National Academy of Sciences of the United States of America* **108**, 21093-
809 21098, doi:10.1073/pnas.1112061109 (2011).
- 810 23 Ding, N., Wu, H., Tao, T. & Peng, E. NEAT1 regulates cell proliferation and
811 apoptosis of ovarian cancer by miR-34a-5p/BCL2. *Onco Targets Ther* **10**,
812 4905-4915, doi:10.2147/OTT.S142446 (2017).
- 813 24 Rokavec, M. *et al.* Corrigendum. IL-6R/STAT3/miR-34a feedback loop
814 promotes EMT-mediated colorectal cancer invasion and metastasis. *J Clin
815 Invest* **125**, 1362, doi:10.1172/JCI81340 (2015).
- 816 25 Su, X. *et al.* TApx63 suppresses metastasis through coordinate regulation of
817 Dicer and miRNAs. *Nature* **467**, 986-990, doi:10.1038/nature09459
818 (2010).
- 819 26 Chang, T. C. *et al.* Widespread microRNA repression by Myc contributes to
820 tumorigenesis. *Nat Genet* **40**, 43-50, doi:10.1038/ng.2007.30 (2008).
- 821 27 Ahn, Y. H. *et al.* ZEB1 drives prometastatic actin cytoskeletal remodeling
822 by downregulating miR-34a expression. *J Clin Invest* **122**, 3170-3183,
823 doi:10.1172/JCI63608 (2012).
- 824 28 Stahlhut, C. & Slack, F. J. Combinatorial Action of MicroRNAs let-7 and
825 miR-34 Effectively Synergizes with Erlotinib to Suppress Non-small Cell
826 Lung Cancer Cell Proliferation. *Cell cycle* **14**, 2171-2180,
827 doi:10.1080/15384101.2014.1003008 (2015).
- 828 29 Yang, P. *et al.* TGF-beta-miR-34a-CCL22 signaling-induced Treg cell
829 recruitment promotes venous metastases of HBV-positive hepatocellular
830 carcinoma. *Cancer Cell* **22**, 291-303, doi:10.1016/j.ccr.2012.07.023
831 (2012).
- 832 30 Wang, X. *et al.* Tumor suppressor miR-34a targets PD-L1 and functions as
833 a potential immunotherapeutic target in acute myeloid leukemia. *Cell
834 Signal* **27**, 443-452, doi:10.1016/j.cellsig.2014.12.003 (2015).

- 835 31 Leveille, N. *et al.* Genome-wide profiling of p53-regulated enhancer RNAs
836 uncovers a subset of enhancers controlled by a lncRNA. *Nat Commun* **6**,
837 6520, doi:10.1038/ncomms7520 (2015).
- 838 32 Rashi-Elkeles, S. *et al.* Parallel profiling of the transcriptome, cistrome,
839 and epigenome in the cellular response to ionizing radiation. *Sci Signal* **7**,
840 rs3, doi:10.1126/scisignal.2005032 (2014).
- 841 33 Hunten, S. *et al.* p53-Regulated Networks of Protein, mRNA, miRNA, and
842 lncRNA Expression Revealed by Integrated Pulsed Stable Isotope Labeling
843 With Amino Acids in Cell Culture (pSILAC) and Next Generation
844 Sequencing (NGS) Analyses. *Mol Cell Proteomics* **14**, 2609-2629,
845 doi:10.1074/mcp.M115.050237 (2015).
- 846 34 Ashouri, A. *et al.* Pan-cancer transcriptomic analysis associates long non-
847 coding RNAs with key mutational driver events. *Nat Commun* **7**, 13197,
848 doi:10.1038/ncomms13197 (2016).
- 849 35 Kim, K. H., Kim, H. J. & Lee, T. R. Epidermal long non-coding RNAs are
850 regulated by ultraviolet irradiation. *Gene* **637**, 196-202,
851 doi:10.1016/j.gene.2017.09.043 (2017).
- 852 36 Consortium, E. P. An integrated encyclopedia of DNA elements in the
853 human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 854 37 Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an
855 alignment-free logistic regression model. *Nucleic acids research* **41**, e74,
856 doi:10.1093/nar/gkt006 (2013).
- 857 38 Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using
858 sequence features and support vector machine. *Nucleic acids research* **35**,
859 W345-349, doi:10.1093/nar/gkm391 (2007).
- 860 39 Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative
861 proteomic analysis of eleven common cell lines reveals ubiquitous but
862 varying expression of most proteins. *Mol Cell Proteomics* **11**, M111
863 014050, doi:10.1074/mcp.M111.014050 (2012).
- 864 40 Tarasov, V. *et al.* Differential regulation of microRNAs by p53 revealed by
865 massively parallel sequencing: miR-34a is a p53 target that induces
866 apoptosis and G1-arrest. *Cell cycle* **6**, 1586-1593,
867 doi:10.4161/cc.6.13.4436 (2007).
- 868 41 Sun, F. *et al.* Downregulation of CCND1 and CDK6 by miR-34a induces cell
869 cycle arrest. *FEBS letters* **582**, 1564-1568,
870 doi:10.1016/j.febslet.2008.03.057 (2008).
- 871 42 Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its
872 antisense RNA. *Nature* **451**, 202-206, doi:10.1038/nature06468 (2008).
- 873 43 Serviss, J. T., Johnsson, P. & Grander, D. An emerging role for long non-
874 coding RNAs in cancer metastasis. *Frontiers in genetics* **5**, 234,
875 doi:10.3389/fgene.2014.00234 (2014).
- 876 44 Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs
877 within cell-cycle promoters. *Nat Genet* **43**, 621-629, doi:10.1038/ng.848
878 (2011).
- 879 45 Ng, S. Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs
880 promote pluripotency and neuronal differentiation by association with
881 chromatin modifiers and transcription factors. *EMBO J* **31**, 522-533,
882 doi:10.1038/emboj.2011.459 (2012).

- 883 46 Ng, S. Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA
884 RMST interacts with SOX2 to regulate neurogenesis. *Molecular cell* **51**,
885 349-359, doi:10.1016/j.molcel.2013.07.017 (2013).
- 886 47 Boque-Sastre, R. *et al.* Head-to-head antisense transcription and R-loop
887 formation promotes transcriptional activation. *Proceedings of the*
888 *National Academy of Sciences of the United States of America* **112**, 5785-
889 5790, doi:10.1073/pnas.1421197112 (2015).
- 890 48 Wang, L. *et al.* A long non-coding RNA targets microRNA miR-34a to
891 regulate colon cancer stem cell asymmetric division. *eLife* **5**,
892 doi:10.7554/eLife.14620 (2016).
- 893 49 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research*
894 **12**, 996-1006, doi:10.1101/gr.229102. Article published online before
895 print in May 2002 (2002).
- 896 50 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI:
897 current status, taxonomic expansion, and functional annotation. *Nucleic*
898 *acids research* **44**, D733-745, doi:10.1093/nar/gkv1189 (2016).
- 899 51 miR34AasRNaproject (2017).
- 900 52 Team, R. C. *R: A Language and Environment for Statistical Computing*,
901 <<https://www.R-project.org/>> (2017).
- 902 53 liftr: Containerize R Markdown Documents v. R package version 0.7
903 (2017).
- 904 54 tidyverse: Easily Install and Load the 'Tidyverse' v. R package version
905 1.2.1 (2017).
- 906 55 rlang: Functions for Base Types and Core R and 'Tidyverse' Features v. R
907 package version 0.1.4 (2017).
- 908 56 magrittr: A Forward-Pipe Operator for R v. R package version 1.5 (2014).
- 909 57 knitr: A General-Purpose Package for Dynamic Report Generation in R v.
910 R package version 1.17 (2017).
- 911 58 rmarkdown: Dynamic Documents for R v. R package version 1.8 (2017).
- 912 59 ggthemes: Extra Themes, Scales and Geoms for 'ggplot2' v. R package
913 version 3.4.0 (2017).
- 914 60 gggenes: Draw Gene Arrow Maps in 'ggplot2' v. R package version
915 0.2.0.9003.
- 916 61 extrafont: Tools for using fonts v. R package version 0.17 (2014).
- 917 62 scales: Scale Functions for Visualization v. R package version 0.5.0 (2017).
- 918 63 gtable: Arrange 'Grobs' in Tables v. R package version 0.2.0 (2016).
- 919 64 A Package for Survival Analysis in S v. version 2.38 (2015).
- 920 65 Holman, J. D., Tabb, D. L. & Mallick, P. Employing ProteoWizard to Convert
921 Raw Mass Spectrometry Data. *Curr Protoc Bioinformatics* **46**, 13 24 11-19,
922 doi:10.1002/0471250953.bi1324s46 (2014).
- 923 66 Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal
924 database search tool for proteomics. *Nat Commun* **5**, 5277,
925 doi:10.1038/ncomms6277 (2014).
- 926 67 Granholm, V. *et al.* Fast and accurate database searches with MS-
927 GF+Percolator. *J Proteome Res* **13**, 890-897, doi:10.1021/pr400937n
928 (2014).
- 929 68 Boekel, J. *et al.* Multi-omic data analysis using Galaxy. *Nature*
930 *biotechnology* **33**, 137-139, doi:10.1038/nbt.3134 (2015).

- 931 69 Amarzguioui, M., Rossi, J. J. & Kim, D. Approaches for chemically
932 synthesized siRNA and vector-mediated RNAi. *FEBS letters* **579**, 5974-
933 5981, doi:10.1016/j.febslet.2005.08.070 (2005).
934 70 Polson, A., Durrett, E. & Reisman, D. A bidirectional promoter reporter
935 vector for the analysis of the p53/WDR79 dual regulatory element.
936 *Plasmid* **66**, 169-179, doi:10.1016/j.plasmid.2011.08.004 (2011).
937 71 Turner, A. M., Ackley, A. M., Matrone, M. A. & Morris, K. V. Characterization
938 of an HIV-targeted transcriptional gene-silencing RNA in primary cells.
939 *Human gene therapy* **23**, 473-483, doi:10.1089/hum.2011.165 (2012).
940 72 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25
941 years of image analysis. *Nature methods* **9**, 671-675 (2012).
942 73 Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N
943 Engl J Med* **375**, 1109-1112, doi:10.1056/NEJMmp1607591 (2016).
944

945 **Figure legend**

946
947 **Figure 1: Characterization of the *IncTAM34a* transcript.** **A)** Architecture of
948 the *miR34a* locus (hg38, RefSeq) including *miR34a* HG, mature *miR34a*, and
949 *IncTAM34a* (*LINC01759*). H3K4me3 ChIP-seq data, indicating the active
950 promoter region, and conservation are also shown. **B)** Semi-quantitative PCR
951 data from the screening of a panel of cancer cell lines. Wild type *TP53* is
952 indicated with +, - indicates null, and +* represents either a non-null *TP53*
953 mutation or wild-type *TP53* with mechanisms present that inhibit its function
954 (e.g. SV40 large T antigen in HEK293T cells). **C)** TCGA correlation analysis.
955 Expression was log2 normalized to the maximum expression value.
956 Nonsynonymous *TP53* mutations are indicated on the top of the plot (cancer
957 type abbreviation definitions and corresponding statistics are in Figure 1-
958 Supplement 1). **D)** 3'-RACE sequencing results and the annotated *IncTAM34a*
959 (*LINC01759*) are shown. **E)** Semi-quantitative PCR results from the primer
960 walk assay (i.e. common reverse primer (exon 2) and forward primers (F10-
961 F15) staggered upstream of *IncTAM34a*'s annotated start site) performed
962 using HEK293T cells (Figure 1-Supplement 2a details primer placement)
963 **F)** Coding potential analysis assessed via the Coding-potential Assessment
964 Tool including *IncTAM34a*, two known non-coding RNAs (*HOTAIR* and *XIST*),
965 and three protein-coding RNAs (β -actin, Tubulin, and *MYC*).

966

967 **Figure 2: TP53-mediated regulation of the *miR34a* locus.** **A)** Evaluating
968 the effects of 24 hours of treatment with 200 ng/ml doxorubicin on *IncTAM34a*
969 and *miR34a* HG in HCT116 and HEK293T cells.* **B)** Monitoring *miR34a* HG
970 and *IncTAM34a* expression levels during 24 hours of doxorubicin treatment

971 in *TP53^{+/+}* and *TP53^{-/-}* HCT116 cells.* **C)** Quantification of luciferase and
972 renilla levels after transfection of HCT116 and HEK293T cells with the p1
973 construct (Figure 2-Supplement 2 contains a schematic representation of the
974 p1 construct).* **D)** HCT116 cells were co-transfected with the p1 construct and
975 shRNA renilla or shRNA control and subsequently treated with increasing
976 doses of doxorubicin. 24 hours post-treatment, cells were harvested and
977 renilla and luciferase levels were measured using QPCR.* *Individual points
978 represent results from independent experiments, error bars show the 95% CI,
979 black horizontal lines represent the mean, and *P* values are shown over long
980 horizontal lines indicating the comparison tested. All experiments in Figure 2
981 were performed in biological triplicate.

982

983 **Figure 3: *lncTAM34a* positively regulates *miR34a* and its associated**
984 **phenotypes. A)** QPCR-mediated quantification of *miR34a* expression in cell
985 lines stably over-expressing *lncTAM34a*.* **B)** Cell cycle analysis comparing
986 stably over-expressing *lncTAM34a* cell lines to the respective mock
987 control.* **C)** Analysis of cellular growth over time in *lncTAM34a* over-
988 expressing PC3 cells. Points represent the median from 3 independent
989 experiments, the colored shadows indicate the 95% confidence interval, and
990 vertical lines show the minimum and maximum values obtained from the three
991 experiments. **D)** Differential phosphorylated polymerase II binding
992 in *lncTAM34a* over-expressing PC3 cells.* *Individual points represent results
993 from independent experiments, error bars show the 95% CI, black horizontal
994 lines represent the mean, and *P* values are shown over long horizontal lines
995 indicating the comparison tested.

996

997 **Figure 4: Survival analysis in TCGA cancers. A)** Kaplan-Meier survival
998 curves comparing the effects of *TP53*-mutated samples (left), low *lncTAM34a*
999 expression (middle) and low *miR34a* expression (right) to control samples in
1000 papillary kidney cancer (results for other cancers in Figure 4-Supplement 1).
1001 Middle and Right panel include only *TP53* wild type patients where RNAseq
1002 data exists. **B)** Correlation analysis between the effects on the 5-year survival
1003 probability of *TP53*-mutated samples, low *lncTAM34a* expression, and low
1004 *miR34a* expression as indicated. For each variable the 5-year survival
1005 probability was compared to the control group (negative values indicate lower
1006 survival, positive values indicate higher survival). Spearman correlation
1007 coefficients are given on top left of each plot. Each dot indicates one cancer
1008 type (see Fig.1c for legend). Boxplots on the bottom summarize the effects for
1009 the parameter on the x-axis, with indication of *P* values, as calculated using
1010 paired Wilcoxon signed rank test. Low expression was defined as *TP53* non-
1011 mutated samples having expression values in the bottom 10th percentile.

1012

1013 **Figure 5: A graphical summary of the proposed *lncTAM34a***
1014 **function.** Stress stimuli, originating in the cytoplasm or nucleus,
1015 activate *TP53* as well as additional factors. These factors then bind to
1016 the *miR34a* promoter and drive baseline transcription levels of the sense and
1017 antisense strands. *lncTAM34a* serves to further increase *miR34a* HG
1018 transcription levels resulting in enrichment of polymerase II at the *miR34a*
1019 promoter and a positive feed-forward loop. *miR34a* HG then, in turn, is spliced
1020 and processed in multiple steps before the mature *miR34a* binds to the RISC

1021 complex allowing it to repress its targets and exert its tumor suppressive
1022 effects.

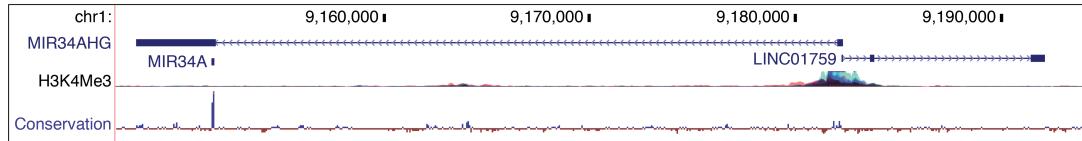
1023 **Figures**

1024

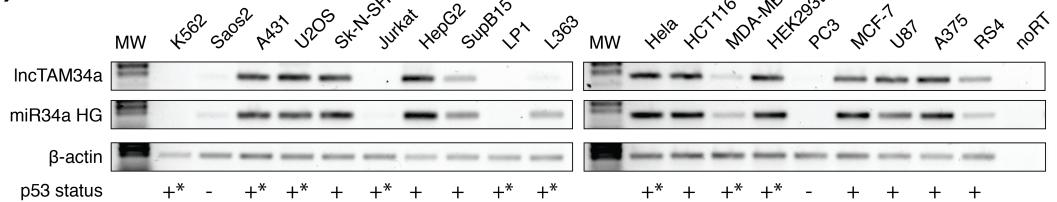
1025

Figure 1

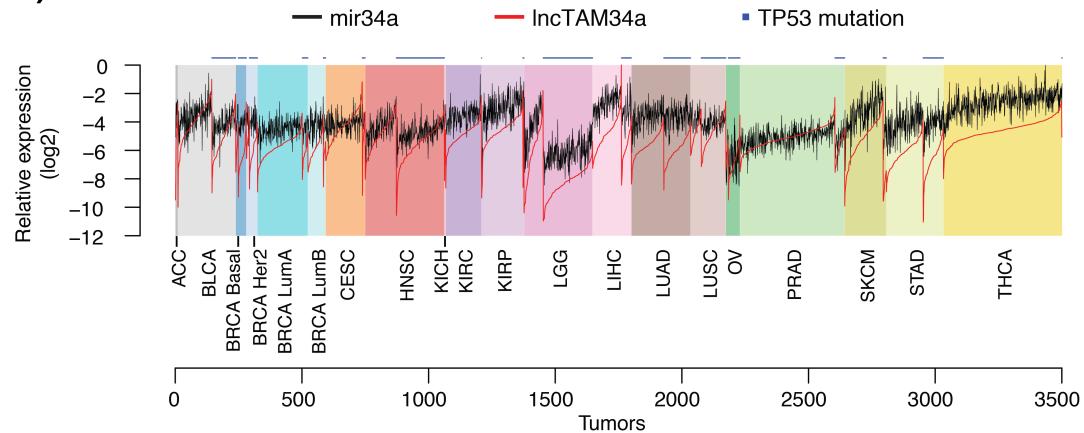
A)



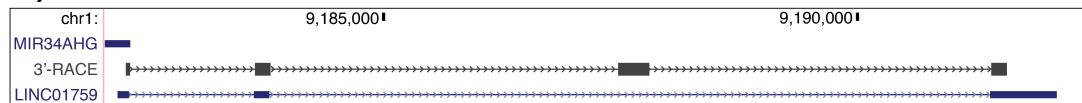
B)



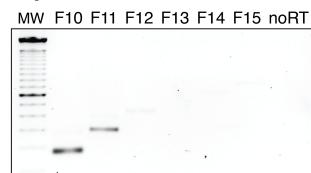
C)



D)



E)

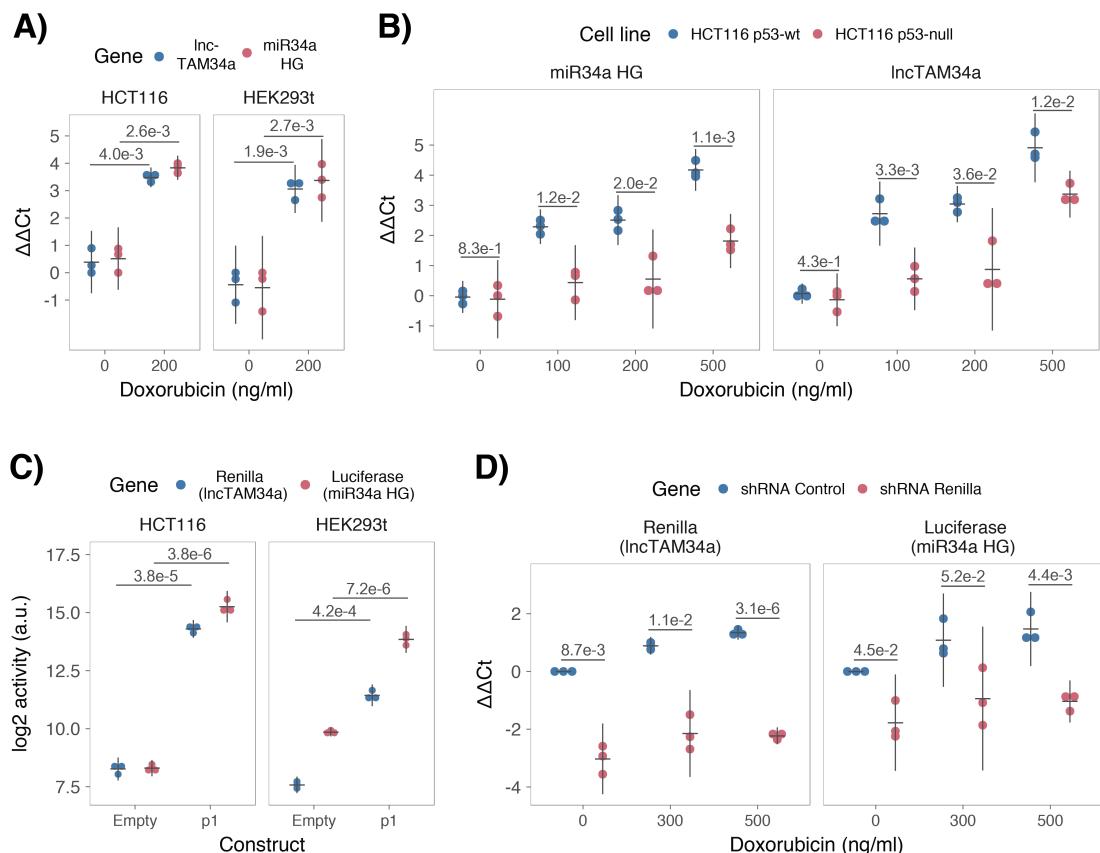


F)

Name	Size	ORF Size	Fickett Score	Hexamer Score	Coding Probability	Coding Label
IncTAM34a	795	285	0.801	0.258	0.301	no
HOTAIR	918	144	0.919	0.268	0.126	no
XIST	19280	411	0.663	-0.128	0.027	no
B-actin	1917	1128	1.350	0.698	1.000	yes
Tubulin	2632	1119	1.277	0.515	1.000	yes
MYC	2345	1320	1.223	0.484	1.000	yes

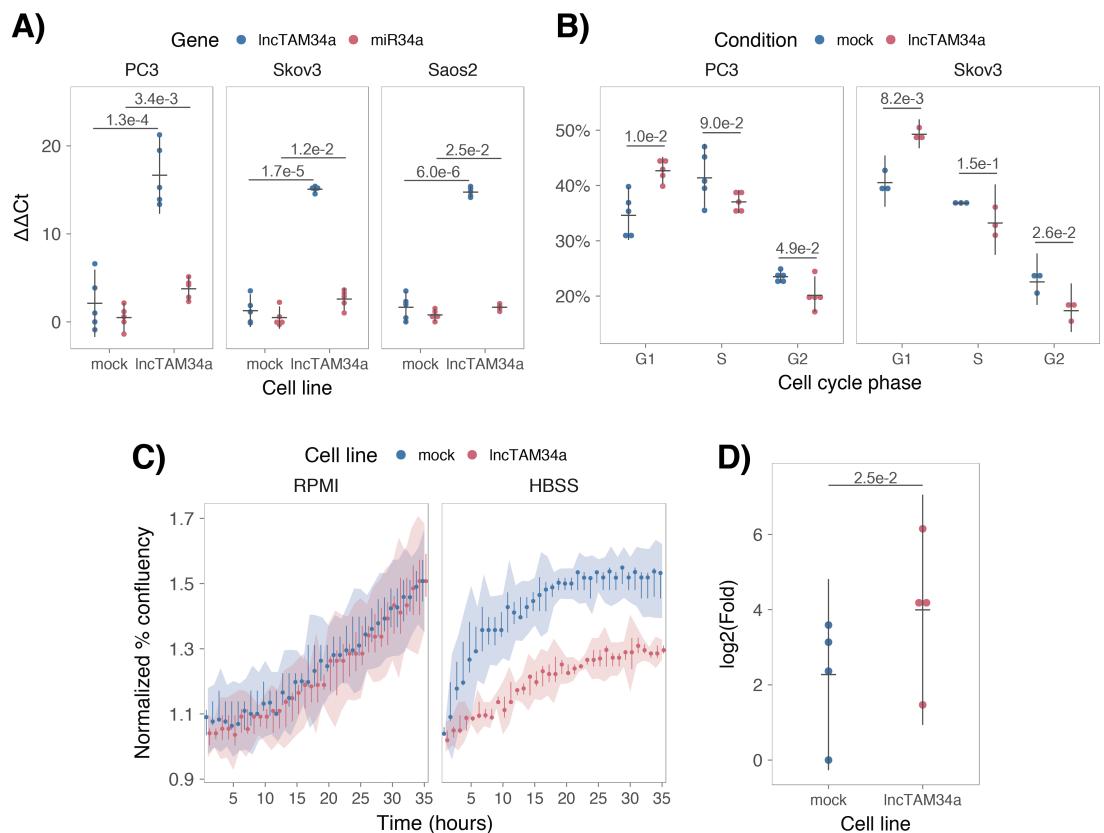
1026

1027 **Figure 2**



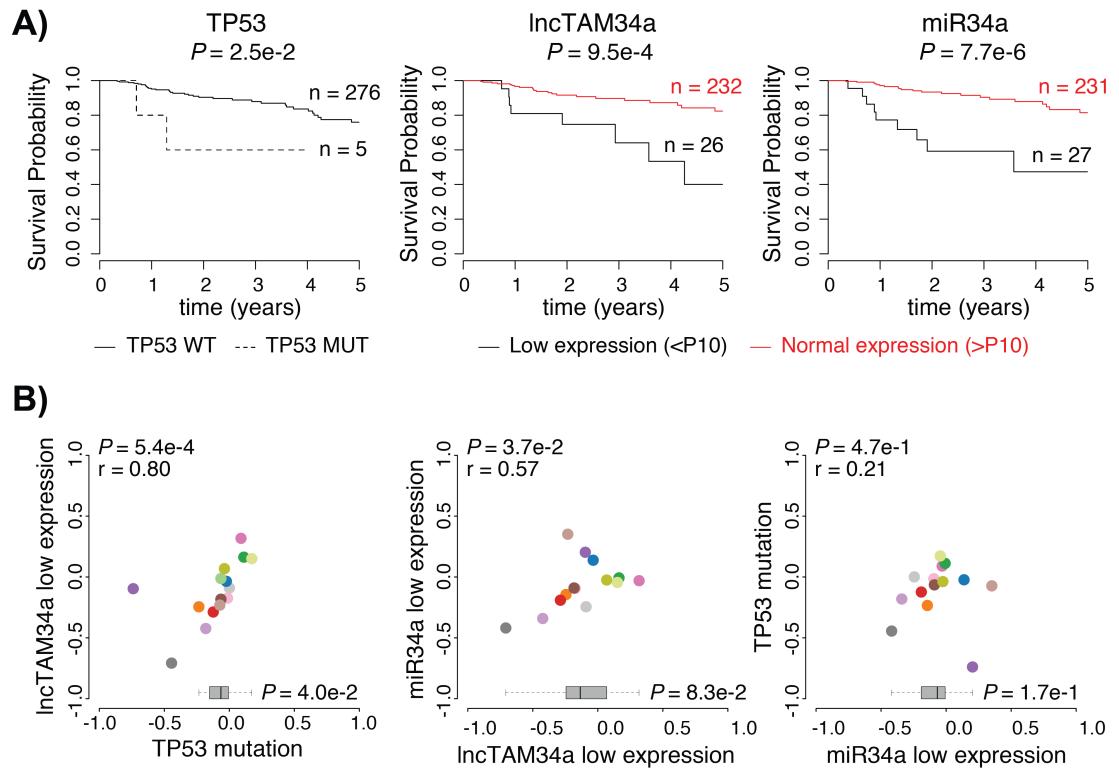
1028

1029 **Figure 3**



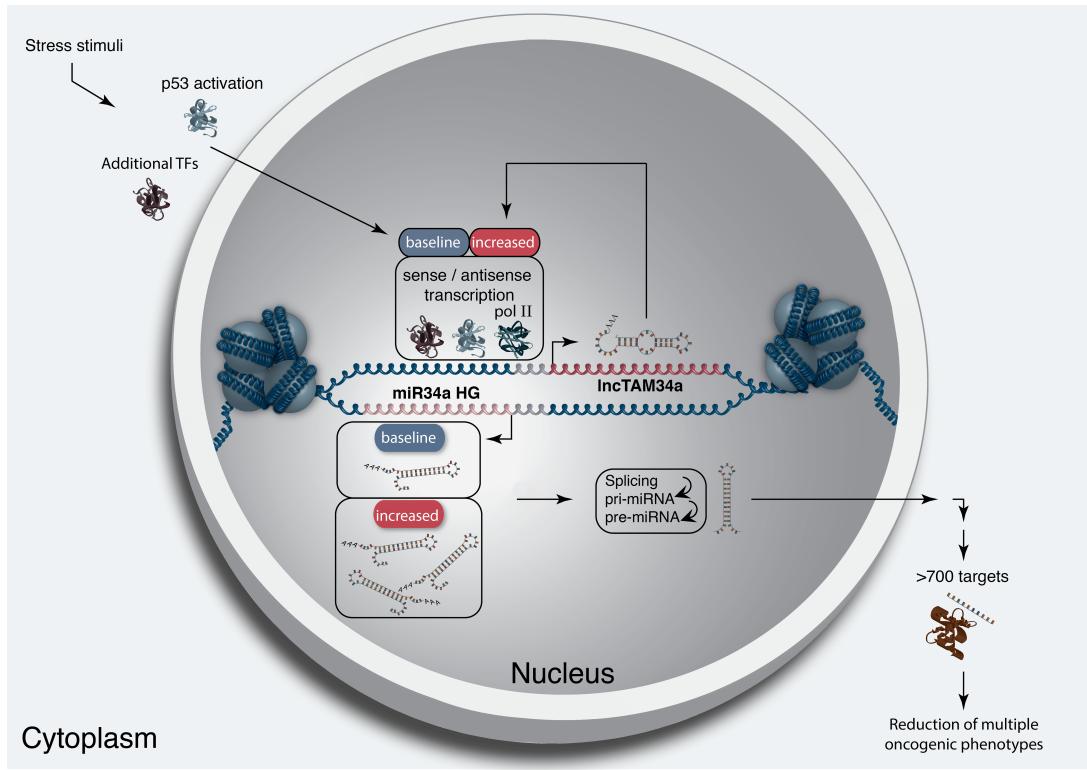
1030

1031 **Figure 4**



1032
1033

1034 **Figure 5**



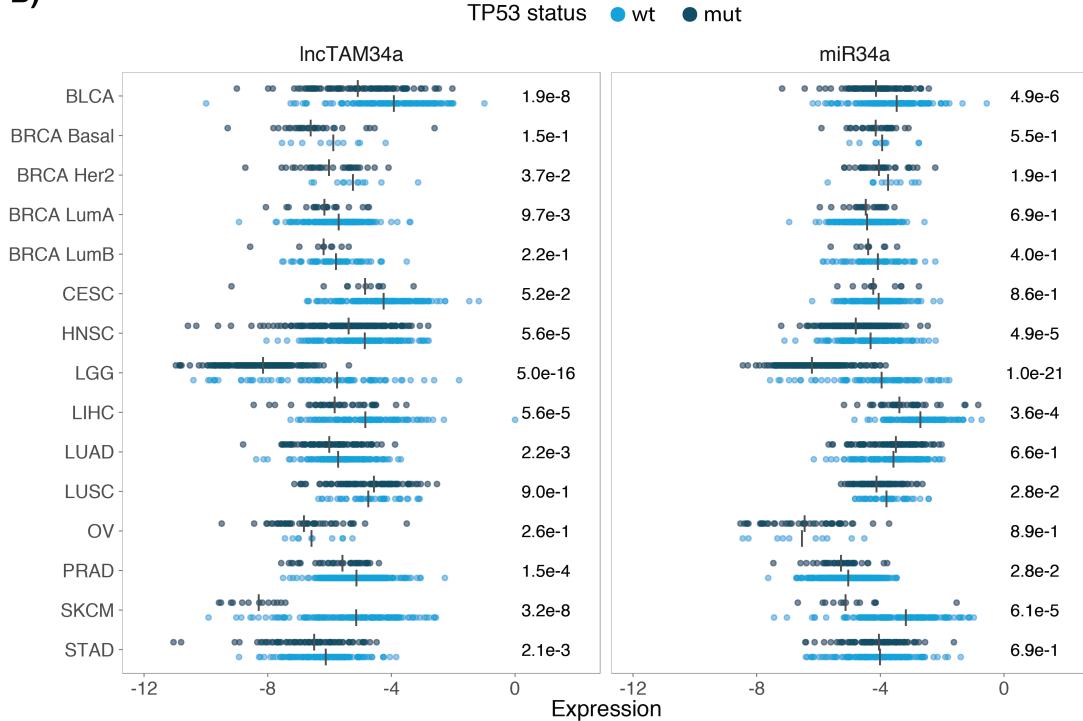
1035

1036 **Supplementary Figures**

A)

cancer	all n	all rho	all p	TP53wt n	TP53wt rho	TP53wt p	TP53mut n	TP53mut rho	TP53mut p
Adrenocortical carcinoma (ACC)	10	0.55	1.04e-01	10	0.55	1.04e-01	NA	NA	NA
Bladder Urothelial Carcinoma (BLCA)	228	0.51	7.89e-17	134	0.45	3.86e-08	94	0.43	1.73e-05
Breast invasive carcinoma (BRCA) Basal	42	0.57	9.54e-05	10	0.62	6.02e-02	32	0.57	7.41e-04
Breast invasive carcinoma (BRCA) Her2	44	0.15	3.39e-01	12	0.22	4.85e-01	32	0.07	7.10e-01
Breast invasive carcinoma (BRCA) LumA	199	0.34	8.22e-07	177	0.34	2.96e-06	22	0.49	2.31e-02
Breast invasive carcinoma (BRCA) LumB	70	0.17	1.57e-01	61	0.15	2.53e-01	9	0.17	6.78e-01
Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC)	156	0.14	8.37e-02	145	0.16	5.45e-02	11	-0.05	9.03e-01
Head and Neck squamous cell carcinoma (HNSC)	313	0.54	8.38e-25	123	0.61	0.00e+00	190	0.45	9.68e-11
Kidney Chromophobe (KICH)	5	0.60	3.50e-01	5	0.60	3.50e-01	NA	NA	NA
Kidney renal clear cell carcinoma (KIRC)	142	0.35	2.06e-05	141	0.34	4.41e-05	NA	NA	NA
Kidney renal papillary cell carcinoma (KIRP)	167	0.45	9.16e-10	163	0.45	2.04e-09	4	0.80	3.33e-01
Brain Lower Grade Glioma (LGG)	271	0.63	9.92e-32	76	0.73	0.00e+00	195	0.39	2.26e-08
Liver hepatocellular carcinoma (LIHC)	153	0.56	3.64e-14	114	0.52	4.18e-09	39	0.45	3.95e-03
Lung adenocarcinoma (LUAD)	234	0.28	1.15e-05	128	0.36	2.87e-05	106	0.23	1.91e-02
Lung squamous cell carcinoma (LUSC)	139	0.23	6.74e-03	42	0.04	7.93e-01	97	0.33	9.91e-04
Ovarian serous cystadenocarcinoma (OV)	56	0.23	8.37e-02	10	0.84	4.46e-03	46	0.15	3.31e-01
Prostate adenocarcinoma (PRAD)	413	0.47	1.33e-23	375	0.46	6.13e-21	38	0.45	4.58e-03
Skin Cutaneous Melanoma (SKCM)	165	0.65	5.43e-21	152	0.61	7.85e-17	13	0.43	1.40e-01
Stomach adenocarcinoma (STAD)	225	0.37	8.23e-09	145	0.37	5.71e-06	80	0.42	1.03e-04
Thyroid carcinoma (THCA)	469	0.46	1.07e-25	467	0.46	4.06e-26	NA	NA	NA

B)



1037
1038

1039 **Figure 1 Supplement 1: TCGA normalized expression levels and correlation analysis statistics.**
1040 A) Spearman's rho and P values (p) from the correlation analysis in Figure 1a between *miR34a* and
1041 *lncTAM34a* expression in *TP53* wild type (wt) and mutated (mut) samples within TCGA cancer types.
1042 NA indicates not applicable, due to a lack of data for the specific group. B) Expression levels of
1043 *miR34a* and *lncTAM34a* in *TP53* wt and nonsynonymous mutation samples. Expression was quantified
1044 by the log2 ratio of expression of the gene to its maximal expression value. Vertical lines indicate the
1045 median. P values are indicated on the right side of each panel and are derived from comparing the
1046 *TP53* wild type samples to the samples with a nonsynonymous mutation using a two-sided Wilcoxon
1047 signed rank test. Only cancers that had at least 5 samples per group were included. In addition, only
1048 samples that were diploid at the *miR34a* locus were used for the analysis to avoid copy number bias.

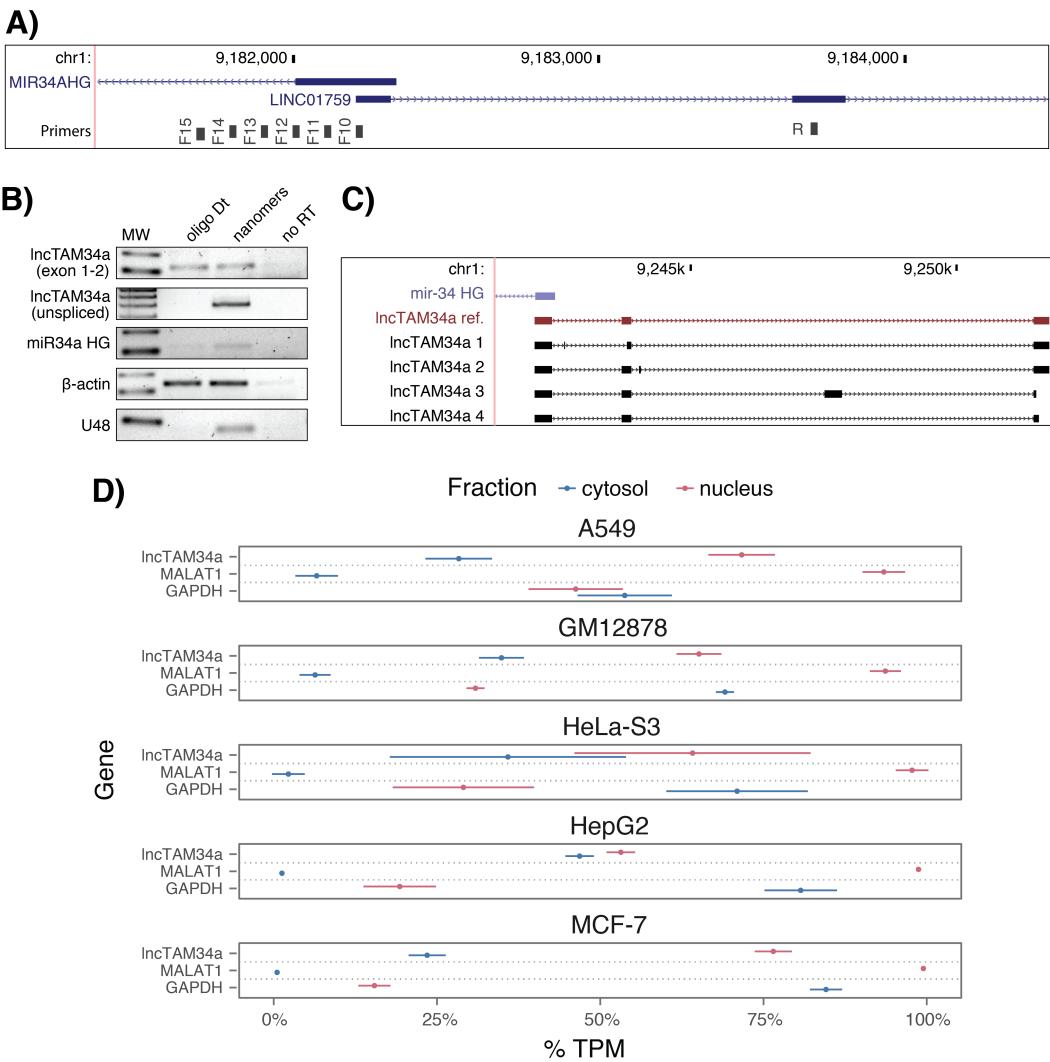
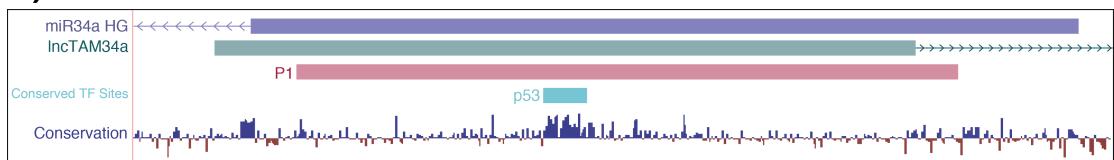
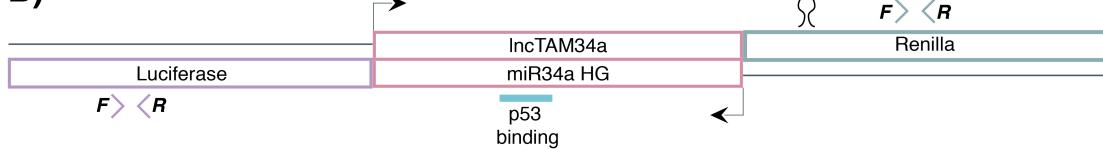


Figure 1 Supplement 2: Molecular characteristics of lncTAM34a. **A)** A schematic representation of the primer placement in the primer walk assay. **B)** Polyadenylation status of spliced and unspliced lncTAM34a in HEK293T cells. **C)** Sequencing results from the analysis of lncTAM34a isoforms in U2OS cells. lncTAM34a ref. refers to the full-length transcript as defined by the 3'-RACE and the primer walk assay. **D)** Analysis of coding potential of the lncTAM34a transcript using the Coding-potential Calculator. **E)** RNAseq data from five fractionated cell lines in the ENCODE project showing the percentage of transcripts per million (TPM) for lncTAM34a. MALAT1 (nuclear localization) and GAPDH (cytoplasmic localization) are included as fractionation controls. Points represent the mean and horizontal lines represent the standard deviation from two biological replicates.

A)

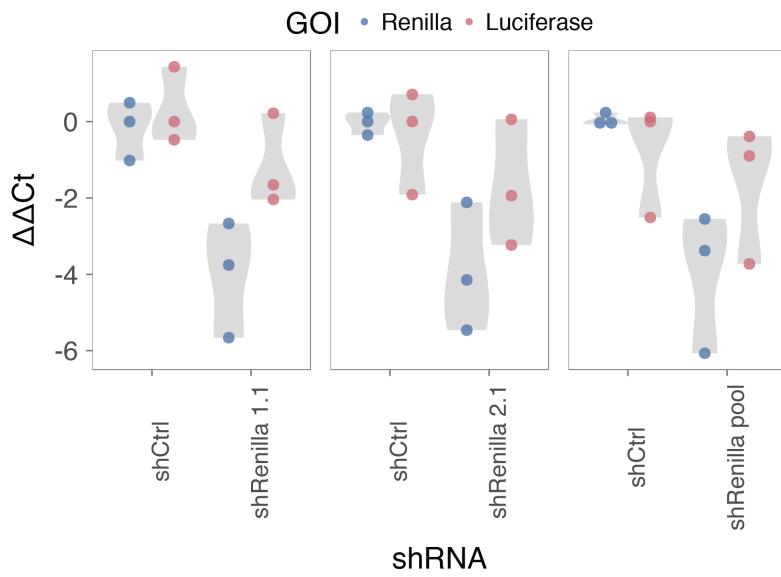


B)

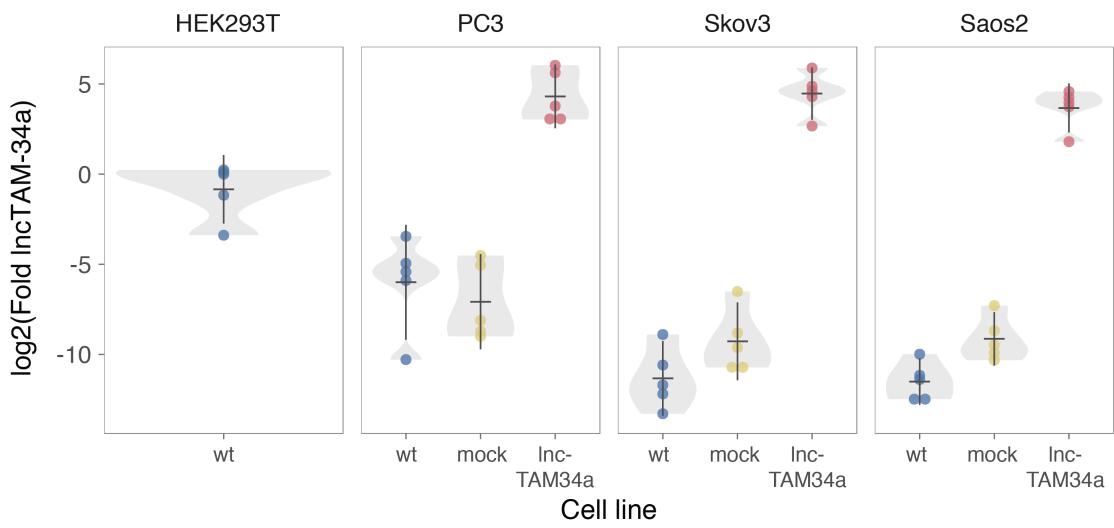


1061
1062
1063
1064
1065
1066

Figure 2 Supplement 1: A schematic representation of the p1 construct. A) A UCSC genome browser illustration indicating the location of the promoter region cloned into the p1 construct including the conserved TP53-binding site. **B)** A representative picture of the p1 construct including forward (F) and reverse (R) primer locations and the renilla shRNA targeting site.



1067
 1068 **Figure 2 Supplement 2: Evaluating the effects of *IncTAM34a* down-regulation.** HEK293T cells
 1069 were co-transfected with the p1 construct and either shRenilla or shControl. Renilla and luciferase
 1070 levels were measured with QPCR 48 hours after transfection. Individual points represent independent
 1071 experiments with the gray shadow indicating the density of the points. The experiment was performed
 1072 in biological triplicate.
 1073



1074
1075
1076
1077
1078

Figure 3 Supplement 1: Physiological relevance of *lncTAM34a* over-expression. Comparison of *lncTAM34a* expression in HEK293T cells (high endogenous *lncTAM34a*), and the wild-type (wt), mock, and *lncTAM34a* over-expressing stable cell lines.

1079
1080
1081
1082
1083

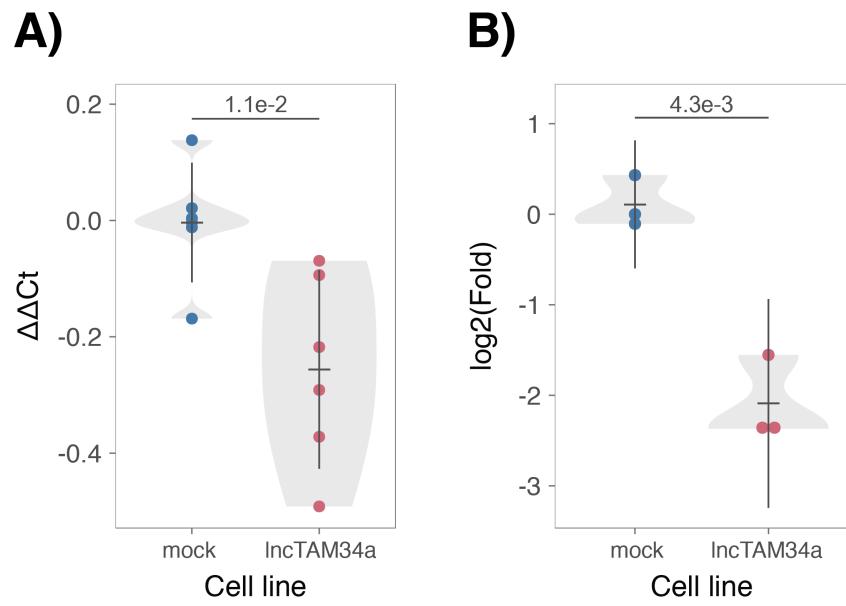
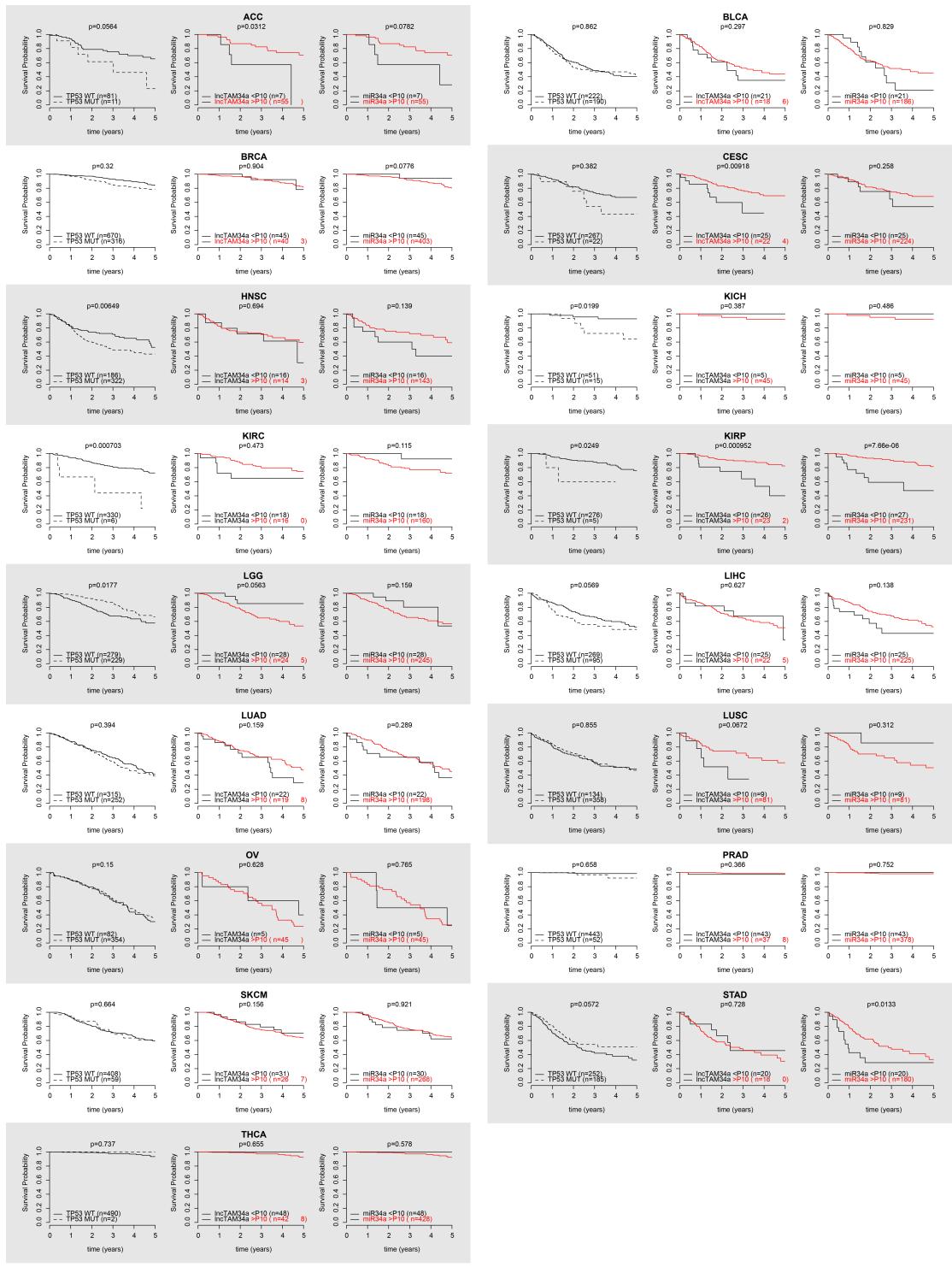


Figure 3 Supplement 2: Effects of *lncTAM34a* over-expression on cyclin D1. CCND1 expression (A) and western blot quantification of protein levels (B) in *lncTAM34a* over-expressing PC3 stable cell lines. Experiments were performed in biological sextuplets (A) or triplicates (B).



1084

— TP53 WT -- TP53 MUT — Low expression (<P10) — Normal expression (>P10)

1085
1086
1087
1088

Figure 4-Supplement 1: Survival analysis in 17 cancers from TCGA. Kaplan-Meier survival curves comparing the survival of *TP53*-mutated samples (left), low *IncTAM34a* expression (middle) and low *miR34a* expression (right) to control samples in 17 cancer types from TCGA. Low expression was defined as *TP53* non-mutated samples having expression values in the bottom 10th percentile.

1089