6mer seed toxicity in tumor suppressive microRNAs

Quan Q. Gao[1,6], William E. Putzbach[1,6], Andrea E. Murmann[1], Siquan Chen[2], Aishe A. Sarshad[3], Johannes M. Peter[4], Elizabeth T. Bartom[5], Markus Hafner[3], and Marcus E. Peter[1,5,*]

[1]Department of Medicine/Division Hematology/Oncology and [5]Department of Biochemistry and Molecular Genetics, Northwestern University, Chicago, IL 60611, USA, [2]Cellular Screening Center, Institute for Genomics & Systems Biology, The University of Chicago, Chicago, IL 60637, USA, [3]Laboratory of Muscle Stem Cells and Gene Regulation, NIAMS, NIH, Bethesda, MD 20892, USA, [4]DigiPen Institute of Technology, Redmond, WA 98052, USA,

[6] Authors share first authorship

Corresponding author: Marcus E. Peter, E-mail: m-peter@northwestern.edu, phone: 312-503-1291; FAX: 312-503-0189.

Keywords: RNAi, miRNAs, strand selection, toxicity, evolution

SUMMARY

Many small interfering (si)RNAs are toxic to cancer cells through a 6mer seed sequence (position 2-7 of the guide strand). A siRNA screen with all 4096 possible 6mer seed sequences in a neutral siRNA backbone revealed a preference for guanine in positions 1 and 2 and a high overall G or C content in the seed of the most toxic siRNAs for four tested human and mouse cell lines. Toxicity of these siRNAs stems from targeting survival genes with C-rich 3'UTRs. The master tumor suppressor miRNA miR-34a-5p is toxic through such a G-rich 6mer seed and is upregulated in cells subjected to genotoxic stress. In the absence of most canonical miRNAs, a modified highly toxic form of miR-320a-3p is induced. An analysis of all mature miRNAs suggests that during evolution most miRNAs evolved to avoid guanine at the 5' end of the 6mer seed sequence of the guide strand. In contrast, for certain tumor suppressive miRNAs the guide strand contains a G-rich toxic 6mer seed, presumably to eliminate cancer cells.

INTRODUCTION

RNA interference (RNAi) is a form of post-transcriptional regulation exerted by 19-21 nt long double stranded RNAs that negatively regulate gene expression at the mRNA level. RNAi-active guide RNAs can come from endogenous siRNAs and micro(mi)RNAs. For a miRNA, the RNAi pathway begins in the nucleus with transcription of a primary miRNA precursor (pri-miRNA) [1]. Pri-miRNAs are first processed by the Drosha/DGCR8 microprocessor complex into pre-microRNAs [2], which are then exported from the nucleus to the cytoplasm by Exportin 5 [3]. Once in the cytoplasm, Dicer processes them further [4, 5] and these mature dsRNA duplexes are then loaded into Argonaute (Ago) proteins to form the RNA-induced silencing complex (RISC) [6]. The sense/passenger strand is ejected/degraded, while the guide strand remains associated with the RISC [7]. Depending on the degree of complementarity between the guide strand and its target, the outcome of RNAi can either be target degradation - most often achieved by siRNAs with full complementarity to its target mRNA [8] - or miRNA-like cleavage-independent silencing, mediated by deadenylation/degradation or translational repression [9]. The latter mechanism can be initiated with as little as six nucleotide base-pairing between a guide RNA's so-called seed sequence (positions 2 to 7) and fully complementary seed matches in the target RNA [10, 11]. This seed-based targeting most often occurs in the 3'UTR of a target mRNA [12, 13].

A number of miRNAs function either as tumor suppressors or as oncogenes [14]. Their cancer specific activities are usually explained by their identified targets, being oncogenes tumor or suppressors, respectively [14]. Examples of targets of tumor-suppressive miRNAs are the oncogenes Bcl-2 for miR-15/16 [15] and c-Myc for miR-34a [16]. While many miRNAs have been reported to have both tumor suppressive and oncogenic activities depending on the cancer context, examples for widely established tumor promoting miRNAs are miR-221/222, miR-21, miR-155, and members of the miR-17~92 cluster, or its paralogues miR-106b~25 and miR-106a~363

[17, 18]. In contrast, two of the major tumor suppressive miRNA families are miR-15/16 and the p53 regulated miR-34a/c and miR-34b [19].

We recently discovered that many si- and shRNAs can kill all tested cancer cell lines through RNAi by targeting the 3'UTRs of critical survival genes [20]. Cancer cells have difficulty developing resistance to this mechanism both *in vitro* and when treated *in vivo* [21]. We reported that a 6mer seed sequence in the toxic siRNAs is sufficient for effective killing [20]. We have now performed a strand specific siRNA screen with a library of individual siRNAs representing all 4096 possible 6mer seed sequences in a neutral RNA duplex. This screen, while based on siRNA biochemistry was not designed to identify targets that are degraded through siRNA mediated slicing activity but to identify toxicity caused by moderately targeting hundreds of genes required for cell survival in a mechanism similar to miRNA-induced silencing.

We report that the most toxic 6mer seeds are G-rich with a G enrichment towards the 5' end targeting survival genes with a high C content in their 3'UTR in a miRNA-like manner. Many tumor suppressive miRNAs such as miR-34a-5p but none of the established oncogenic miRNAs contain G-rich 6mer seeds and most of miR-34a-5p's toxicity comes from its 6mer seed sequence. Mature miRNAs from older and more conserved miRNAs contain less toxic seeds. We demonstrate that for most miRNAs the more abundant mature form corresponds to the arm that contains the less toxic seed. In contrast, for major tumor suppressive miRNAs, the mature miRNA is derived from the arm that harbors the more toxic seed. Our data allow us to conclude that while most miRNAs have evolved to avoid targeting survival and housekeeping genes, certain tumor suppressive miRNAs function to kill cancer cells through a toxic G-rich 6mer seed targeting the 3'UTR of survival genes.

RESULTS

Identifying the most toxic 6mer seeds

To test how certain 6mer seeds present in the guide strand of a siRNA affect cancer cell survival, we recently designed a neutral 19mer oligonucleotide scaffold with two nucleotide 3' overhangs, and we demonstrated that modifying a siRNA strand at positions 1 and 2 by 2'-O-methylation (OMe) completely blocks its loading into the RISC [22]. Different 6mer sequences can be inserted at positions 2-7 of the guide strand with the designated passenger strand modified by OMe (two red Xs in **Fig. 1a**). Transfection efficiency and conditions were optimized for each cell line used. To determine the general rules of seed-based toxicity, we individually transfected 4096 siRNAs with all possible 6mer seed sequences in this 19mer scaffold into two human, HeyA8 (ovarian cancer) and H460 (lung cancer), and two mouse cell lines M565 (liver cancer) and 3LL (lung cancer). This allowed us to rank all 4096 6mer seeds according to their toxicity (**Fig. 1b**, **Supplementary Table 1,** and **6merdb.org**). The congruence between the results of the two human cell lines and the two human and two mouse cell lines was quite high (r=0.68 and 0.65, respectively, **Fig. 1c**) suggesting that many

siRNAs were toxic through a mechanism independent of cancer origin and species. Toxicity was caused by the different 6mer seeds in the *guide* strand. A siRNA duplex highly toxic to all cell lines (#2733, HeyA8 cell viability = 1.4%) strongly inhibited cell growth and reduced cell viability of HeyA8 cells only when the passenger strand was modified by the OMe modification. Toxicity was completely blocked when the guide strand was modified (**Fig. S1a**). The toxicity was due to RNAi as knockdown of AGO2 abolished the toxicity of two of the most toxic siRNAs (**Fig. S1b**).

We previously reported that the CD95 ligand (CD95L) coding region (CDS) is enriched in sequences that when converted into si- or shRNAs are toxic to cancer cells [20] and that the CD95L mRNA itself is toxic to cells [23]. We now report a substantial correlation between the most toxic CD95L-derived shRNAs and the toxicity of their predicted 6mer seed (**Fig. 1d**) suggesting the CD95L-derived si/shRNAs kill cancer cells through 6mer seed toxicity. Consistent with this assumption we found that the 6mer seeds of 4 previously tested siRNAs derived from CD95L [24] in this screen were about as toxic as the full length siRNAs, with $siL3^{Seed}$ being the most toxic followed by $siL2^{Seed}$ and less or no toxicity associated with $siL4^{Seed}$ and $siL1^{Seed}$ (**Fig. 1b, Fig. 1c**). Our recent analysis suggested that the toxic si/shRNAs act like artificial miRNAs by targeting the 3'UTR of mRNAs [20].

6mer seeds enriched in G at the 5'end are most toxic to human and mouse cell lines

We noticed that the 6mer seeds in siL3 and siL2 have a higher G content than the ones in siL4 and siL1 (**Fig. 1b**). By analyzing the screen results of all four cell lines (**Fig. S2**), we found that a high G content of the seed correlated better with toxicity than a high C content. Almost no toxicity was found with seeds with a high A content. To test the effect of nucleotide content on toxicity directly, we retested the 19 seed duplexes with the highest content (>80%) for each of the four nucleotides in the four cell lines (**Fig. 2**). The reanalysis also allowed us to determine the reproducibility of the results obtained in the large screens (which for technical reasons had to be performed in three sets). All data on the three cell lines were highly reproducible especially for the most toxic seeds (**Fig. S3a**). When the data on the four cell lines were compared, it became apparent that in all cell lines, the G-rich seeds were by far the most toxic followed by the C-rich, U-rich, and A-rich seeds (**Fig. 2a**). This indicates it is mostly the G content that determines toxicity.

Most genome-wide siRNA libraries designed to study functions of individual genes are highly underrepresented in G and C to increase RNAi specificity [25] (**Fig. S3b**, left panel). In contrast, our complete set of 6mer seed duplexes exhibits no nucleotide composition bias, allowing us to test the contributions of all four nucleotides in each of the 6 seed positions (**Fig. S3b**, right panel). To determine the nucleotide content of the most toxic seeds, we averaged the nucleotide content at each of the 6 positions of either the 200 most or 200 least toxic seed duplexes for each of the two human and the two mouse cells (**Fig. 2b**, **Supplementary Table 1**). In all four cell lines, we found that a high G content towards the 5' end of the seed and a C in

133 position 6 was most toxic (**Fig. 2b** and **Fig. S3c**). In contrast, non-toxic seeds were A and U-rich especially
134 when positioned at the 5' end. While the rules of toxicity that emerged are almost identical between human
135 and mouse, we noticed a subtle difference. C-rich 6mer seeds are slightly more toxic to mouse than to human
136 cells. This can be seen in the nucleotide composition of the most toxic seeds (**Fig. 2b**), a slightly higher
137 toxicity of the most C rich seeds in the siRNA retest (**Fig. 2a**), and this can also be explored at 6merdb.org.
138

139 **siRNAs with toxic seeds target housekeeping genes enriched in C-rich sequences in their 3'UTR**

140 We previously showed that si- and shRNAs are toxic through 6mer seed toxicity preferentially targeting
141 hundreds of genes critical for cell survival [20]. We had developed a Toxicity Index (TI), a simple tool to predict
142 the most toxic seeds based on the ratio of putative seed match occurrences in the 3'UTR of set of survival
143 genes (SGs) versus a set of genes not involved in cell survival (nonSGs). We now compared the TI with our
144 experimentally determined 6mer seed toxicity in the three cell lines screened (**Fig. S4**) and found a significant
145 correlation between these two types of analyses, further supporting the mechanism of toxicity. Knowing that
146 seed sequences rich in G are most toxic suggested the targeted genes carry C-rich seed matches in their 3'UTR.
147 To be most stringent, we used a list of the 20 6mer seed containing siRNAs that were most toxic to both
148 HeyA8 and H460 cells (**Supplementary Table 2**). The G richness towards the 5' end of the 6mers in these
149 toxic seeds and the 5' A/U richness of the nontoxic seeds was even more pronounced than in the top/bottom
150 200 most toxic seeds (**Fig. 3a**). We scored for the occurrence of seed matches to the 20 seeds in each group in
151 the 3'UTR, the CDS and the 5'UTR of a set of 938 critical SGs similar to one recently described [20] and an
152 expression-matched set of 938 nonSGs. We found a significantly higher count ratio of toxic versus nontoxic
153 seed matches in the 3'UTR of SGs when compared to nonSGs (**Fig. 3b**, right panel). Consistent with a
154 miRNA-like function no such enrichment was found when the CDS was analyzed (**Fig. 3b**, center panel). An
155 inverse ratio of sequences complementary to the 6mer seeds of unknown significance was found in the 5'UTR
156 (**Fig. 3b**, left panel). This result was very similar when we scored for seed matches to the 100 seeds most toxic
157 and least toxic to both human cell lines (**Fig. S5a**). An enrichment of the exact seed matches in 3'UTRs was
158 consistent with the overall higher C content of 3'UTRs of SGs when compared to nonSGs (different peak
159 maxima in **Fig. 3c**). A metaplot analysis of the 500 bases upstream and downstream of the translational start
160 and stop site of all human genes showed that as expected the 3'UTR was enriched in A and U (**Fig. 3d**, top).
161 Interestingly, SGs had a lower A/U content in a region ~150-500 bases into the 3'UTR than expression
162 matched nonSGs (**Fig. 3d**, green horizontal bar, center and bottom). To determine where survival genes are
163 being targeted by toxic seed containing siRNAs we again performed a metaplot analysis this time plotting the
164 locations of seed matches to the 20 6mer seeds that were most and least toxic to both human cell lines (**Fig. 3e**,
165 an analysis with the 100 most/least toxic seeds is provided in **Fig. S5b**). When analyzing all human coding
166 genes we found the reverse complements of the most toxic seeds to be highly enriched at the beginning of the

167 3'UTR whereas the reverse complements of the least toxic seeds were underrepresented in this region (**Fig.**
168 **3e**, top). This effect was not due to a much higher G/C or lower A/U content in this region (**Fig. 3d**, top). A
169 comparison of the location of these seed matches in the SGs and in expression matched nonSGs confirmed
170 this general trend, however, two differences between SGs and nonSGs became apparent: 1) nonSGs have
171 more nontoxic seed matches ~150-500 bases into the 3'UTR (**Fig. 3e**, bottom, green horizontal bar) maybe due
172 to the higher A/U content of this region (**Fig. 3d**, center and bottom). 2) SGs contain a small stretch at
173 positions 42-65 into the 3'UTR (**Fig. 3e** and **Fig. S5b**, center, red horizontal bar) that is enriched in seed
174 matches for the most toxic seeds. This region in SGs seems to be a preferential target site for siRNAs carrying
175 toxic G-rich seed sequences.
176

177 **Tumor suppressive miR-34a-5p kills cancer cells through its toxic 6mer seed**

178 The toxic siRNAs kill cancer cells through 6mer seed toxicity by a mechanism reminiscent of the function of
179 miRNAs. To test whether actual miRNAs could kill cancer cells with the help of toxic 6mer seeds, we
180 analyzed the seed toxicity determined in our screen for all known ~2600 mature miRNAs expressed as either
181 the 3p or 5p arm (**6merdb.org**). While none of the 6mer seeds present in the predominant arm (guide strand)
182 of the most oncogenic miRNAs (miR-221/222, miR-21, miR-155, the miR-17~92, miR-106b~25, and miR-
183 106a~36 clusters) were toxic (reduced viability >50%, stippled line in **Fig. 4a**), two of the major tumor
184 suppressive miRNA families, miR-15/16 and p53 regulated miR-34a/c and miR-34b contained toxic seeds in
185 the guide strand (**Fig. 4a**). This suggested these two families were killing cancer cells through toxic 6mer
186 seeds. Interestingly, two other major tumor suppressive families, let-7 and miR-200, did not contain toxic G-
187 rich seeds in their guide strand, suggesting they may be tumor suppressive through other mechanisms, such as
188 inducing and maintaining cell differentiation [26].

189 When transfecting the pre-miRs of miR-34a-5p, miR-15a-5p and let-7a-5p into HeyA8 cells, the potency
190 of these three miRNAs to reduce cell growth mimicked the toxicity of their 6mer seed containing siRNAs (**Fig.**
191 **4b** and **4a**). This suggested that a large part of their toxicity comes from the composition of the seed position
192 2-7. The most toxic seed in a major tumor suppressive miRNA was present in miR-34a5p/34c-5p, a master
193 regulator of tumor suppression [27]. We directly compared the toxicity of pre-miR-34a-5p and its toxic seed in
194 the neutral scaffold with blocked passenger strand (si34a-5p^{Seed}) in the same assays (**Fig. 4c**). Strikingly, the
195 toxicity evoked by these two RNA species (assessed by growth inhibition and DNA fragmentation) was
196 similar. Cells showed the typical morphology we found in cells dying from toxic siRNAs (**Fig. 4d**) [20, 24, 28]. To
197 determine the contribution of the 6mer seed sequence of miR-34a-5p to its toxicity and the mode of cell death,
198 we performed a RNA-Seq analysis on HeyA8 cells transfected with either miR-34a-5p or si34a-5p^{Seed} (**Fig. 4e**,
199 top) (**Supplementary Table 3**). The vast majority of genes were significantly up- and downregulated by both
200 RNA duplexes (**Fig. 4e**, bottom). While miR-34a-5p targeted a subset of genes not affected by miR-34a-5p^{Seed},

201 the majority of genes (>78%) were downregulated >1.5 fold by both the premiR and the 6mer seed duplex
202 (**Fig. 4f**, left). A Sylamer analysis is a unbiased approach allowing to identify which seed matches are enriched
203 in the 3'UTRs of downregulated genes from RNAseq data [29]. In this analysis both duplexes caused similar and
204 highly effective downregulation of the mRNAs that carry a 6mer seed match (**Fig. 4g**). When the Sylamer
205 analysis was performed with either 7mer or 8mer seeds, enrichment of seed matches was much less significant
206 (**Fig. S6a**) suggesting that most of the targeting by both RNAs only required a 6mer seed.
207 Consistent with this activity, targeting by both RNA duplexes resulted in a very similar reduction of
208 survival genes (**Fig. 4h**). The genes downregulated by both the premiR and the 6mer seed construct were
209 highly enriched in genes involved in regulation of cell cycle, cell division, DNA repair, and nucleosome
210 assembly (**Fig. 4f**, right). These GO terms were very similar to the ones we found enriched in downregulated
211 genes in cells dying after transfection with CD95R/L-derived si/shRNAs containing toxic 6mer seeds [20]. In
212 contrast, no such GO terms were found enriched when the same analysis was performed with the upregulated
213 genes as control (**Fig. S6b**). While both miR-34-5p and si34a-5p^{Seed} caused the most significant
214 downregulation of genes carrying 8mers in their 3'UTR (**Fig. S6c**), only the most highly downregulated genes
215 that carry the shared 6mer seed match were grouped in a number of GO terms that are consistent with 6mer
216 seed toxicity as previously reported [20] and barely any GO terms were shared among genes that contained 7 or
217 8mer seed matches (**Fig. S6d**). All these data suggest that miR-34a-5p kills cancer cells using its toxic 6mer
218 seed. While optimal miRNA targeting requires at least a 7mer seed interaction and also involves nucleotides at
219 positions 13-16 of the miRNA [30], the cell death inducing activity of this tumor suppressive miRNA may only
220 require the 6mer seed.
221

222 **Toxic 6mer seed toxicity is shaping the miRNA repertoire**

223 Toxic 6mer seeds may be a driving force in miRNA evolution, whereby toxic seed sequences are either
224 selected against - because they contribute to cell toxicity - or are preserved to operate as tumor suppressors.
225 Based on the composition of toxic 6mer seeds and the enrichment of corresponding seed matches in survival
226 genes, we could now ask whether and when miRNAs that contain toxic G-rich sequences in positions 2-7 of
227 their seeds evolved. When comparing all mature miRNA arms annotated in TargetScan Human 7, we noticed
228 that miRNAs in highly conserved miRNA seed families contained 6mer seed sequences that were much less
229 toxic in our screen than seeds in poorly conserved miRNAs (**Fig. 5a,** left panel and **Fig. S7a**). Weakly
230 conserved miRNA seed families would be expected to be younger in evolutionary age than highly conserved
231 ones. Consistent with this assumption we found that the 6mer seeds of younger miRNAs (<10 million years
232 old) were more likely to be toxic to cells than the ones of older miRNAs (>800 million years old) [31] (**Fig. 5a**,
233 right panel and **Fig. S7b**). Most importantly, when comparing miRNAs of different ages, it became apparent
234 that seeds of miRNAs over the last 800 million years were gradually depleted of G beginning at the 5' end and

235 eventually also affecting positions 3-5 until the oldest ones, where A and U had replaced G as the most
236 abundant nucleotide in all six positions (**Fig. 5b**). These analyses indicated that most highly conserved
237 miRNAs avoid G in potentially toxic seed positions. Interestingly, the most toxic seed sequences were found
238 in miRtrons (**Fig. 5c** and **Fig. S7c**), miRNAs that are derived by splicing short introns [32].

239 miRNAs are expressed as pre-miRs and usually only one major species of mature miRNA (either the 5p or
240 the 3p arm) is significantly expressed in cells produced from one of the two strands of the premiR stem [33].
241 Consistent with the assumption that cells cannot tolerate toxic 6mer seeds, we now tested across 780 miRNAs
242 which have been shown to give rise to both a 3p and a 5p arm whether the more highly expressed arm contains
243 a seed of lower toxicity than the lesser expressed arm (**Fig. 5d**). We ranked the miRNAs according to the ratio
244 of the 6mer seed toxicity associated with the guide arm to the lesser-expressed arm (**Supplementary Table 4**).
245 When we labeled the major tumor suppressive and oncogenic miRNAs, we noticed the highly expressed arm
246 of most of the oncogenic miRNAs contained a 6mer seed that was not toxic in our screen (**Fig. 5d**, green
247 dots). In contrast, for most of the tumor suppressive miRNAs, the dominant arm contained a seed much more
248 toxic than the lesser arm (**Fig. 5d**, red dots). The overall difference in ratio between the two groups of
249 miRNAs was highly significant. A more detailed analysis of these data revealed that the three oncogenic
250 miRNAs with the highest ratio in toxicity between their arms, miR-363, miR-92a-2, and miR-25, were almost
251 exclusively expressed as the non-toxic 3p form (**Fig. 5e**, top). In contrast, the dominant arm of the three tumor
252 suppressive miRNAs, miR-34a, miR-34c, and miR-449b contained the most toxic seed sequence (**Fig. 5e**,
253 bottom). Interestingly, miR-449b has the same seed sequence as miR-34a and has been suggested to act as a
254 backup miRNA for miR-34a [34]. These data are consistent with most tumor suppressive miRNAs using 6mer
255 seed toxicity to kill cancer cells and suggest that this mechanism developed over hundreds of millions of
256 years.
257

258 **Genotoxic drugs upregulate toxic 6mer seed containing miRNAs**

259 Our data showing that miR-34a-5p contains a toxic 6mer seed, along with miR-34a being upregulated after
260 genotoxic stress [19], led us to wonder whether miR-34a-5p would contribute to cell death induced by genotoxic
261 drugs and whether this type of cell death had similarities to the death observed in cells dying from toxic 6mer
262 seed containing si/shRNAs. This would be consistent with the observation that many genotoxic drugs induce
263 multiple cell death pathways [35, 36, 37, 38, 39]. To compare cell death induced by different genotoxic agents with that
264 of toxic si/shRNAs, we treated the p53 wild-type ovarian cancer cell line HeyA8 with doxorubicin (Doxo),
265 carboplatin (Carbo), or etoposide (Eto) and performed a RNA-Seq analysis. Drug concentrations were chosen
266 so that after 80 hrs, treatment would slow down cell growth and induce signs of stress without major cell death
267 occurring to capture changes that could be causing cell death rather than being the result of it (**Fig. S8a** and
268 data not shown). The morphological changes in the cells treated with the drugs were very similar to the ones

seen in cells treated with si/shRNAs (**Fig. S8b**), and similar to reported morphologies of cells treated with genotoxic drugs [40, 41].

The ranked lists of downregulated RNAs isolated from HeyA8 cells treated with the three drugs were subjected to a gene set enrichment analysis (GSEA) to determine whether survival genes were enriched in the downregulated genes (**Fig. S9a**). There was strong enrichment of downregulated survival genes towards the top of the ranked list. 102 of the survival genes were downregulated in cells treated with any of the three drugs (**Fig. S9b**). In a DAVID gene ontology analysis, these genes were strongly enriched in clusters involved in chromosome segregation, DNA replication, cell cycle regulation, and mitosis, typical for 6mer seed toxicity induced cell death (**Supplementary Table 5**). We quantified 30 of the 102 survival genes in HeyA8 cells treated with Doxo at different time points using an arrayed quantitative PCR (**Fig. S9c**). 24 of the 30 genes' mRNAs were significantly downregulated as early as 7 hours after treatment with no further reduction beyond 15 hrs after treatment, suggesting that their repression was the cause of cell death rather than a consequence. A Metascape analysis of all RNA-Seq data of downregulated RNAs in response to the toxic siL3, si34a-5p^{Seed}, miR-34a-5p, and the three genotoxic drugs suggested a common mode of action (**Fig. S9d**). The GO clusters that were most significantly downregulated in all data sets were again related to DNA repair, cell cycle, and mitosis as described before [24].

To test whether treatment of cells with genotoxic drugs results in loading the RISC with toxic miRNAs, HeyA8 cells were treated with Doxo for 0, 20, 40 and 80 hrs and all 4 Ago proteins were pulled down using a GW182 peptide [42]. Interestingly, while the amount of AGO2 pulled down was the same at all time points, the amount of bound miRNA-sized RNAs substantially increased with longer treatment times (**Fig. 6a**). This was most likely the result of an overall increase in total small RNAs in the treated cells (**Fig. 6b**). Alternatively, this could also be a result of cells dividing more slowly and a stable RISC. After Doxo treatment, >99% of the Ago-bound small RNAs were miRNAs (data not shown). miR-34a/b/c-5p bound to Ago proteins were upregulated at all time points (**Fig. S10a**). To determine the contribution of miR-34a-5p and other miRNAs to the toxicity seen in cells exposed to the genotoxic drugs, we treated Drosha k.o. cells - devoid of most canonical miRNAs [43] - with the three genotoxic drugs (**Fig. S11**). These cells were hypersensitive to the toxicity induced by any of the three drugs. We attributed this response to the absence of most canonical miRNAs that protect cells from toxic RNAi active sequences [20]. This result also suggested involvement of small RNAs that do not require Drosha for processing. As expected, the composition of small RNAs bound to Ago proteins dramatically varied between wild-type and Drosha k.o. cells (**Fig. 6c**). In the absence of most canonical miRNAs, miR-320a-3p, which was previously shown not to require Drosha for its biogenesis [43], represented more than 86% of all Ago-bound miRNAs. Similar to HeyA8 cells (see **Fig. S10a**), Ago-bound miR-34a-5p was upregulated in wild-type but not in Drosha k.o. HCT116 cells upon Doxo treatment (**Fig. S10b**, right). Interestingly, the average 6mer seed toxicity of all Ago-bound miRNAs >1.5 fold upregulated in

303 HCT116 wt cells was significantly higher than the ones >1.5 fold downregulated in cells treated with Doxo
304 (**Fig. 6d**, left). While in the Drosha k.o. cells, a number of nontoxic miRNAs were downregulated, the only
305 miRNA that was upregulated in the RISC (1.49 fold) was miR-320a-3p (**Fig. 6d**, right). However, upon closer
306 inspection it became clear that this form of miR-320a-3p was shortened by two nucleotides at the 5' end. This
307 resulted in the shift of the 6mer seed into a G-rich sequence (**Fig. 6d**, right), converting a moderately toxic
308 miRNA (average viability = 49.2%) into a highly toxic one (average viability = 9.3%, **Fig. 1b**). To test this
309 predicted increase in toxicity experimentally, we transfected HeyA8 cells with either the authentic pre-miR-
310 320a-3p or a miR-320a-3p^{Seed} duplex that corresponded to the shifted Ago-bound miR-320a-3p sequence
311 (miR-320a-3p^{Ago}) (**Fig. 6e**). While pre-miR-320a-3p was not toxic, miR-320a-3p^{Ago} completely blocked the
312 growth of the cells. Toxicity of miR-320a-3p^{Ago} was established in four human and mouse cell lines (**Fig. 6f**).
313 These data suggested that in the absence of other miRNAs that could kill cells through 6mer seed toxicity,
314 miR-320a-3p (and possibly other small RNAs) may represent an alternative mechanism that ensures that
315 genotoxic stressors can kill cells with defective miRNA processing often observed in cancer [44, 45]. In summary,
316 our data suggest that certain tumor suppressive miRNAs, such as miR-34a-5p and miR-320a-3p exert their
317 tumor suppressive activities by carrying toxic 6mer seed sequences that can kill cancer cells by targeting
318 survival genes in C-rich regions close to the start of their 3'UTR. This activity may contribute to the cell death
319 induced by genotoxic drugs.
320

321 **DISCUSSION**

322

323 We previously discovered a fundamental cell type- and species-independent form of toxicity that is evoked by
324 the 6mer seed sequence in si-/shRNAs that function similar to miRNAs [20]. We have now performed an siRNA
325 screen that effectively tested the miRNA activities of all 4096 different 6mer seed sequences. Performing the
326 screen in four cell lines (two human and two mouse) ensured that the results were relatively independent of
327 species or cell type specific transcriptomes. The screen has discovered the rules underlying this seed toxicity
328 and allows prediction of the 6mer seed toxicity for any siRNA, shRNA, miRNA with a known 6mer seed
329 (https://6merdb.org).

330 Based on this screen, the toxicity of a number of tumor suppressive miRNAs could be predicted solely on
331 the basis of their 6mer seed sequences. The enrichment of G in the first 2-3 positions of the most toxic seeds is
332 consistent with the way Ago proteins scan mRNAs as targets. This involves mainly the first few nucleotides
333 (positions 1-3) of the seed [46]. miR-34a-5p contains two Gs in positions 1 and 2 of its 6mer seed. While miR-
334 34a-5p is considered a master tumor suppressive miRNA, no single target has been identified to be responsible
335 for this activity. Over 700 targets implicated in cancer cell proliferation, survival, and resistance to therapy
336 have been described [16]. Our data now suggest that miR-34a-5p uses 6mer seed toxicity to target hundreds of

337 housekeeping genes. Our data provide the means to rationally design new artificial miRNAs as anti-cancer
338 reagents that attack networks of survival genes. In humans, miR-34a is highly expressed in many tissues.
339 Consistent with our data that delivering siRNAs with toxic 6mer seeds to mice are not toxic to normal cells [21]
340 miR-34a exhibits low toxicity to normal cells *in vitro* and *in vivo* [47]. miR-34a (MRX34) became the first
341 miRNA to be tested in a phase I clinical trial of unresectable primary liver cancer [27,48]. The study was recently
342 terminated and reported immune-related adverse effects in several individuals. It was suggested that these
343 adverse effects may have been caused by either a reaction to the liposome-based carrier or the use of double-
344 stranded RNA [16]. In addition, they may be due to an undesired gene modulation by miR-34a itself defined by
345 sequences outside the 6mer seed [16]. Our data suggest that miR-34a exerts toxicity mostly through the 6mer
346 seed of its 5p arm and that its 700 known targets may be part of the network of survival genes that are targeted.
347 The comparison of the RNA-Seq data of cells treated with either miR-34a-5p or si34-5p^{Seed} now allows to
348 determine whether these two activities can be separated.

349 Our data provide evidence that genotoxic drugs kill cancer cells, at least in part, by triggering the toxic
350 6mer seed mechanism. Exposure of cancer cells to such drugs resulted in upregulation of tumor suppressive
351 miRNAs, most prominently of the p53 regulated miR-34 family [19]. While one report demonstrated that
352 inhibiting miR-34a rendered cancer cells more resistant to cell death induced by genotoxic stress [49], another
353 one found no effect of knocking out miR-34a on the sensitivity of HCT116 or MCF-7 cells to Doxo [50]. This
354 mechanism may be highly redundant and may involve many miRNAs. Our analysis of Ago-bound miRNAs in
355 Drosha k.o. cells suggest that in the absence of miR-34a, the noncanonical miR-320a-3p which was recently
356 also found to be p53 regulated [51] may act as a backup miRNA that can still respond to genotoxic stress in case
357 the amounts of other miRNAs are reduced, for instance in cases of mutations in miRNA biogenesis associated
358 genes frequently found in human cancers [44]. In addition, recent data suggest that other toxic small RNAs can
359 also be taken up by the RISC and negatively regulate cell growth through their toxic 6mer seed (data not
360 shown).

361 It was shown before that miRNAs overall avoid seed sequences that target the 3‘UTR of
362 survival/housekeeping genes [52,53]. Survival genes therefore are depleted in seed matches for the most abundant
363 miRNAs in a cell. That also means 3‘UTRs of survival genes must be enriched in sequences not targeted by
364 the seeds present in most miRNAs. Our combined data now suggest it is these sequences that toxic siRNAs
365 and tumor suppressive miRNAs with toxic 6mer seeds are targeting. Our analyses also suggest that most
366 miRNAs have evolved over the last 800 million years by gradually depleting G in their seeds beginning at the
367 5' end. In addition, the most abundant miRNAs have evolved to use the arm with the lower 6mer seed toxicity
368 as the active guide strand, presumably to avoid killing cells. Only in a minority of tumor-suppressive miRNAs
369 does the dominant guide strand contain a toxic seed. By ranking miRNAs according to whether they express

370 the arm with the seed of higher toxicity, it is now possible to identify novel tumor suppressive miRNAs
371 (see 6merdb.org).

372 In summary, we have determined the rules of RNAi targeting by toxic 6mer seeds. These rules allowed us
373 to predict with some confidence which si/shRNAs or miRNAs have the potential to kill cells through their
374 toxic 6mer seed. Toxic miRNAs seem to be involved in killing cancer cells in response to genotoxic drugs.
375 Toxic 6mer seeds are present in a number of tumor-suppressive miRNAs that can kill cancer cells. Our data
376 allow new insights into the evolution of miRNAs and provide evidence that 6mer seed toxicity is shaping the
377 miRNA repertoire.
378

379 **METHODS**

380 **Reagents, cell lines and antibodies**

381 HeyA8 (RRID:CVCL_8878) and H460 (ATCC HTB-177) cells were cultured in RPMI1640 medium (Cellgro
382 Cat#10-040) supplemented with 10% FBS (Sigma Cat#14009C) and 1% L-Glutamine (Corning Cat#25-005).
383 3LL cells (ATCC CRL-1642) were cultured in DMEM medium (Gibco Cat#12430054) supplemented with
384 10% FBS and 1% L-Glutamine. Mouse hepatocellular carcinoma cells M565 were previously described [54] and
385 cultured in DMEM/F12 (Gibco Cat#11330) supplemented with 10% FBS, 1% L-Glutamine and ITS (Corning
386 #25-800-CR). HCT116 parental (Cat#HC19023, RRID:CVCL_0291) and the Drosha k.o. clone (clone #40,
387 Cat#HC19020) were purchased from Korean Collection for Type Cultures (KCTC). Both HCT116 cell lines
388 were cultured in McCoy's 5A medium (ATCC, Cat#30-2007) supplemented with 10% FBS and 1% L-
389 Glutamine. All cell lines were authenticated by STR profiling. Anti-Argonaute-2 antibody (cat#ab186733)
390 was purchased from Abcam, anti-β-actin antibody from Santa Cruz (#sc-47778), and secondary antibody for
391 Western blot was Goat anti-rabbit, IgG-HRP from Southern Biotech (#SB-4030-05). Etoposide (Cat#BML-
392 GR307-0100) was purchased from Enzo Life Sciences, propidium iodide (#P4864) doxorubicin (Cat#D1515)
393 and carboplatin (Cat#C2538) were from Sigma-Aldrich.
394

395 **siRNA screens and cell viability assay**

396 The non-toxic siRNA backbone used in the 4096 screen was designed as previously described [22]. Briefly, the
397 siNT2 sequence was used as a starting point and four positions in the center of siNT2 were replaced with the
398 complementary nucleotides in order to remove any identity between the backbone siRNA and the toxic siL3
399 while retaining the same GC content. Two 2'-O-methylation groups were added to positions 1 and 2 of the
400 passenger strand to prevent loading into the RISC. The 6mer seed region (position 2-7 on the guide strand)
401 was then replaced with one of the 4096 possible seeds. Transfection efficiency was optimized for each of the
402 four cell lines individually. For HeyA8 and M565 cells, the reverse transfection and the quantification of
403 cellular ATP content in the 4096 seed duplexes screen was done as previously described [22]. For 3LL and H460

cells screens were done in a very similar manner except for the amount of RNAiMAX and the number of cells plated. For 3LL cells, 150 cells per well were added in a 384 well plate and 9.3 μl of Lipofectamine RNAiMax was mixed with 990.7 μl of Opti-MEM. For H460 cells, 420 cells per well in a 384 well plate was added and 7.3 μl of Lipofectamine RNAiMax was mixed with 993.7 μl of Opti-MEM.

Transfection with short oligonucleotides

For IncuCyte experiments, HeyA8 cells were plated in 50 μl antibiotic free medium in a 96 well plate at 1000 cells/well, and 50 μl transfection mix with 0.1 μl RNAiMAX and siRNAs or miRNA precursors were added during the plating. For the AGO2 knockdown experiment, 100,000 cells/well HeyA8 cells were reverse-transfected in six-well plates with either non-targeting (Dharmacon, cat#D-001810-10-05) or an AGO2 targeting siRNA SMARTpool (Dharmacon, cat#L004639-00- 005) at 25 nM. 1 μl RNAiMAX per well was used for HeyA8 cells. Twenty-four hours after transfection with the SMARTpools, cells were reversed-transfected in a 96-well plate with siNT2, si2733, or si2733 (see **Supplementary Table 1**) at 10 nM and monitored in the IncuCyte Zoom. 48 hours after transfection, cells were lysed in RIPA buffer for Western blot analysis as previously described [22].

All custom siRNA oligonucleotides were ordered from integrated DNA technologies (IDT) and annealed according to the manufacturer's instructions. In addition to the 4096 siRNAs of the screen the following siRNA sequences were used:

siNT1 sense: rUrGrGrUrUrUrArCrArUrGrUrCrGrArCrUrArATT;

siNT1 antisense: rUrUrArGrUrCrGrArCrArUrGrUrArArArCrCrAAA;

siNT2 sense: rUrGrGrUrUrUrArCrArUrGrUrUrGrUrGrUrGrATT;

siNT2 antisense: rUrCrArCrArCrArArArCrArUrGrUrArArArCrCrAAA;

siL3 sense: rGrCrCrCrUrUrCrArArUrUrArCrCrCrArUrArUTT;

siL3 antisense: rArUrArUrGrGrGrUrArArUrUrGrArArGrGrGrCAA;

si-miR-34a-5p^{Seed} sense: mUmGrGrUrUrUrArCrArUrGrUrArCrUrGrCrCrATT;

si-miR-34a-5p^{Seed} antisense: rUrGrGrCrArGrUrArCrArUrGrUrArArArCrCrAAA;

miR-320a-3p^{Ago} sense: mCmGrCrCrCrUrCrUrCrArArCrCrCrArGrCrUrUTT

miR-320a-3p^{Ago} antisense: rArArGrCrUrGrGrGrUrUrGrArGrArGrGrGrCrGAA.

The following miRNA precursors and negative controls were used: hsa-miR-34a-5p (Ambion, Cat. No# PM11030), hsa-let-7a-5p (Ambion, Cat. No# PM10050), hsa-miR-320a-3p (Ambion, Cat. No# PM11621), hsa-miR-15a-5p (Ambion, Cat. No# PM10235), and miRNA precursor negative control #1 (Ambion, Cat. No# AM17110).

Monitoring growth over time by IncuCyte and cell death assays

Cells were seeded between 1000 and 3,000 per well in a 96-well plate in triplicates. The plate was then scanned using the IncuCyte ZOOM live-cell imaging system (Essen BioScience). Images were captured every six hours using a 10× objective. Cell confluence was calculated using the IncuCyte ZOOM software (version 2015A). For treatment with genotoxic drugs HeyA8 cells were seeded at 750 cells/well and HCT116 cells were seeded at 3000 cells/well in 96-well plate and treated with one of the three genotoxic drugs (carboplatin, doxorubicin, or etoposide) at various concentrations at the time of plating. Solvent treated (0.025% DMSO in medium) cells were used as control for etoposide. Medium treated cells were used as control for carboplatin and doxorubicin. To assess cell viability, treated cells were subjected to a quantification of nuclear fragmentation or ATP content both as previously described [20, 22].

RNA-Seq analysis

For RNA-Seq data in **Fig. 4e**, 50,000 cells/well HeyA8 cells were reversed transfected in 6-well plates with 10 nM of either pre-miR-34a-5p or si-miR-34a-5p^{Seed} with their respective controls. The transfection mix was replaced 24 hours after transfection. Cells were lysed after 48 hours after transfection using Qiazol. For the RNA-Seq data in **Fig. S9**, HeyA8 cells were seeded at 50,000 cells per well in a 6-well plate and treated with three genotoxic drugs in duplicate: carboplatin (25 μg/ml), doxorubicin (50 ng/ml), and etoposide (500 nM). Medium treated cells were used as control for carboplatin and doxorubicin treated cells. Solvent control treated cells (0.025% DMSO in medium) were used as control for etoposide. Cells were lysed after 80 hours drug incubation using Qiazol. Total RNA was isolated using the miRNeasy Mini Kit (Qiagen, Cat.No# 74004) following the manufacturer's instructions. An on-column digestion step using the RNAse-free DNAse Set (Qiagen, Cat.No# 79254) was included for all RNA-Seq samples. The RNA libraries were prepared and sequenced as previously described [20] at the Genomics Core facility at the University of Chicago. Reads were aligned to the hg38 version of the human genome, using either Tophat v2.1.0 (RNA-Seq data in **Fig. S9**) or STAR v2.5.2 (RNA-Seq data in **Fig. 4e**). In either case, aligned reads were associated with genes using HTSeq v0.6.1, and the UCSC hg38 transcriptome annotation from iGenomes. Differentially expressed genes were identified using the edgeR R package.

Arrayed real-time PCR

The top 30 most downregulated survival genes shared among HeyA8 cells treated with carboplatin, doxorubicin, and etoposide based on the RNA-Seq analysis were selected for a kinetics analysis using real-time PCR. To prepare the RNAs for the kinetics analysis, 75,000/plate HeyA8 cells were plated in 15 cm plate. Twenty-four hours after plating, one plate of HeyA8 cells were lysed in QIAzol as the control sample. The rest of the plates were treated with 50 ng/ml doxorubicin for 7 hrs, 14.5 hrs, and 21 hrs respectively

471 before being lysed in QIAzol. To perform the arrayed real-time PCR, 200 ng total RNA per sample was
472 used as the input to make cDNA using the high-capacity cDNA reverse Transcription Kit (Applied
473 Biosystems #4368814). For TaqMan Low Density Array (TLDA) profiling, custom-designed 384-well TLDA
474 cards (Applied Biosystems, Cat. No#4346799) were selected and used according to the manufacturer's
475 protocols. For each sample, 20 μl cDNA was mixed with 80 μl water and 100 μl TaqMan Universal PCR
476 Master Mix (Applied Biosystems, Cat. No#4304437). A total volume of 100μl of each sample is loaded into
477 the 8 loading ports on the TLDA card (2 ports for each sample, 4 samples total on one card). The qPCR assays
478 used to detect the 30 genes on the TLDA card are as follows: HIST1H2AI (Hs00361878_s1), CENPA
479 (Hs00156455_m1), HJURP (Hs00251144_m1), FAM72D (Hs00416746_m1), CCNA2 (Hs00996788_m1),
480 KIF20A (Hs00993573_m1), PRC1 (Hs01597839_m1), KIF15 (Hs01085295_m1), BUB1B
481 (Hs01084828_m1), SCD (Hs01682761_m1), AURKA (Hs01582072_m1), NUF2 (Hs00230097_m1), NCAPH
482 (Hs01010752_m1), SPC24 (Hs00699347_m1), KIF11 (Hs00189698_m1), TTK (Hs01009870_m1), PLK4
483 (Hs00179514_m1), AURKB (Hs00945858_g1), CEP55 (Hs01070181_m1), HMGCS1 (Hs00940429_m1),
484 TOP2A (Hs01032137_m1), KIF23 (Hs00370852_m1), INCENP (Hs00934447_m1), CDK1
485 (Hs00938777_m1), HIST2H2BE (Hs00269023_s1), KNL1 (Hs00538241_m1), NCAPD2 (Hs00274505_m1),
486 RACGAP1 (Hs01100049_mH), SPAG5 (Hs00197708_m1), KNTC1 (Hs00938554_m1). GAPDH
487 (Hs99999905_m1) was used as the endogenous control. qPCR assay for individual gene was done in technical
488 triplicates on each TLDA cards. Statistically analysis was performed using Student's t test.
489

490 **Ago affinity peptide purification**

491 The FLAG-GST-T6B WT and mutant peptides was expressed and purified as described [42]. Briefly, constructs
492 were expressed in BL21-Gold(DE3)pLysS competent cells (Agilent). Bacteria, induced with 1 mM Isopropyl
493 β-D-1-thiogalactopyranoside (IPTG), were grown in 1 L overnight at 18ºC to OD 0.6. The bacteria were
494 sedimented at 4000*g* for 15 min and resuspended in 25 ml GST-A buffer (1 mM 4-(2-aminoethyl)
495 benzenesulfonyl fluoride hydrochloride (AEBSF), 1 mM DTT in PBS) supplemented with 1 mg/mL lysozyme
496 (Sigma). Samples were sonicated three times for 3 min at 100% amplitude (Sonics, VCX130) and cleared by
497 centrifugation at 20,000*g* for 20 min. The lysate was loaded onto a column containing 2 ml of bead volume
498 glutathione Sepharose beads (Sigma) and washed two times with GST-A buffer. The GST-tagged protein was
499 eluted in 10 ml of GST-B buffer (20 mM Tris, pH 8.0, and 10 mM glutathione in PBS). The peptide was
500 concentrated using Amicon Ultra-15 Centrifugal Filter Unit (Millipore) and desalted using Zeba Spin
501 Desalting Columns (ThermoFisher).
502

503 **Ago pull down and small RNA-seq**

504 HeyA8 (5-7 x 10^6), HCT116 wild-type (1.2-1.6 x 10^8) or Drosha k.o. (4.8-6.3 x 10^7) cells treated with
505 doxorubicin were lysed in NP40 lysis buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% (v/v)
506 NP40, supplemented with phosphatase inhibitors) on ice for 15 minutes. The lysate was sonicated 3 times for
507 30 s at 60% amplitude (Sonics, VCX130) and cleared by centrifugation at 12,000*g* for 20 minutes. AGO1-4
508 were pulled down by using 500 µg of Flag-GST-T6B peptide [42] and with 60 µl anti-Flag M2 magnetic beads
509 (Sigma-Aldrich) for 2 hrs at 4°C. The pull-down was washed 3 times in NP40 lysis buffer. During the last
510 wash, 10% of beads were removed and incubated at 95°C for 5 minutes in 2x SDS-PAGE sample buffer.
511 Samples were run on a 4-12% SDS-PAGE and transferred to nitrocellulose membrane. The pull-down
512 efficiency was determined by immunoblotting against AGO2 (Abcam 32381). To the remaining beads 500 µl
513 TRIzol reagent were added and the RNA extracted according to the manufacturer's instructions. The RNA
514 pellet was diluted in 20 µl of water. The sample was split and half of the sample was dephosphorylated with
515 0.5 U/µl of CIP alkaline phosphatase at 37°C for 15 min and subsequently radiolabeled with 0.5 µCi γ-^{32}P-
516 ATP and 1 U/µl of T4 PNK kinase for 20 min at 37°C. The AGO1-4 interacting RNAs were visualized on a
517 15% urea-PAGE. The remaining RNA was taken through a small RNA library preparation as previously
518 described [55]. Briefly, RNA was ligated with 3' adenylated adapters and separated on a 15% denaturing urea-
519 PAGE. The RNA corresponding to insert size of 19-35 nt was eluted from the gel, ethanol precipitated
520 followed by 5' adapter ligation. The samples were separated on a 12% Urea-PAGE and extracted from the gel.
521 Reverse transcription was performed using Superscript III reverse transcriptase and the cDNA amplified by
522 PCR. The cDNA was sequenced on Illumina HiSeq 3000.
523 Adapter sequences:
524 Adapter 1 – NNTGACTGTGGAATTCTCGGGTGCCAAGG;
525 Adapter 2 – NNACACTCTGGAATTCTCGGGTGCCAAGG,
526 Adapter 3 – NNACAGAGTGGAATTCTCGGGTGCCAAGG,
527 Adapter 4 – NNGCGATATGGAATTCTCGGGTGCCAAGG,
528 Adapter 47 – NNTCTGTGTGGAATTCTCGGGTGCCAAGG,
529 Adapter 48 – NNCAGCATTGGAATTCTCGGGTGCCAAGG,
530 Adapter 49 – NNATAGTATGGAATTCTCGGGTGCCAAGG,
531 Adapter 50 – NNTCATAGTGGAATTCTCGGGTGCCAAGG.
532 RT primer sequence: GCCTTGGCACCCGAGAATTCCA;
533 PCR primer sequences:
534 CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA.
535
536 **Data analyses**

537 GSEA was performed using the GSEA software version 3.0 from the Broad Institute downloaded from
538 https://software.broadinstitute.org/gsea/. A ranked list was generated by sorting genes according the Log_{10}(fold
539 downregulation). The Pre-ranked function was used to perform GSEA using the ranked list. 1000
540 permutations were used. Default settings were used. The ~1800 survival genes and ~420 non-survival genes
541 defined previously [20] were used as custom gene sets. Default settings were used.

542 The list of survival genes and expression matched non-survival genes were generated by taking the
543 survival and expression matched non-survival genes used previously (*20*) and retaining only the 938 genes in
544 each group of expression matched survival and non-survival genes with an average expression across all RNA
545 seq datasets above 1000 RPMs (see **Supplementary Table 6**).

546 Sylamer analysis [29] was used to find enrichment of small word motifs in the 3'UTRs of genes enriched in
547 those that are most downregulated. A list of 3'UTRs was generated by eliminating repetitive regions as
548 described previously [20] to use for the Sylamer analysis. Default settings were used. Bonferroni adjusted p-
549 values were calculated by multiplying the unadjusted p-values by the number of permutations for each length
550 of word searched for.

551 The GO enrichment analyses shown in **Fig. 4f** and **Fig. S6b** were performed using the GOrilla GO
552 analysis tool at http://cbl-gorilla.cs.technion.ac.il using default setting using different p-value cut-offs for each
553 analysis. GO analysis in **Supplementary Table S5** was done using DAVID 6.8 (https://david.ncifcrf.gov)
554 using default settings. GO analyses across multiple data sets were performed using the software available on
555 www.Metascape.org and default running parameters.

556 Density plots showing the contribution of the four nucleotides G, C, A and U at each of the 6mer seed
557 positions were generated using the Weblogo tool at http://weblogo.berkeley.edu/logo.cgi using the frequency
558 plot setting.

559 Venn diagrams were generated using http://bioinformatics.psb.ugent.be/webtools/Venn/ using default
560 settings.

561 The scatter plot in **Fig. 4e** was generated using R package ggplot2. 10875 genes with RPM > 1 (average
562 RPM of the 8 RNAseq samples) and adjusted p-value < 0.05 were included. 3696 genes were significantly
563 upregulated in both mir-34a-5p and si34a-5p^{Seed} treated samples. 4207 genes were significantly
564 downregulated in both mir-34a-5p and si34a-5p^{Seed} treated samples. 713 genes were only downregulated and
565 792 genes were only upregulated in si34a-5p^{Seed} treated samples. 730 genes were only downregulated and 737
566 genes were only upregulated in mir-34a-5p treated samples. 193 genes out of the total 10875 genes were
567 omitted in the graph as the range for X and Y axes were set as -3 to 3.

568

569 **Identification the 6mer seeds most and least toxic to human cancer cells.**

570 To identify the 20 and 100 most and least toxic seeds to both human cell lines all 4096 seeds were ranked
571 for each cell line from highest to lowest toxicity. The 20 seeds with the highest toxicity to both HeyA8 and
572 H460 cells were found in the top 46 most toxic seeds to in both cells and the 20 seeds shared to be least toxic
573 were found in the bottom 149 seeds in each ranked group. The 100 most and least toxic seeds for both cell
574 lines were identified in the same way and all groups of seeds are shown in **Supplementary Table 2**.
575

576 **Metaplots of 6-mer seed match locations**

577 3pUTR sequences were downloaded from Ensembl Biomart. In order to reduce redundancy in the sequences,
578 a single longest 3pUTR (and associated transcript) was chosen to represent each gene. A custom perl script
579 (makeSeedBed.pl) was written to identify exact matches to all seeds (reverse complement) in all sequences,
580 and to output the coordinates of those matches in bed file format. A custom R script (plotBedMetaPlot.R) was
581 written that uses the GenomicRanges (https://doi.org/10.1371/journal.pcbi.1003118) and Sushi
582 (https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtu379) R packages to calculate the coverage of seeds
583 across all sequences in a given set, and to create a plot of that coverage. The custom scripts and the input data
584 are available in the cloud-based computational reproducibility platform Code Ocean at
585 https://codeocean.com/capsule/9a3eb292-6e89-44f0-b9b0-bdd779f97516/code .
586

587 **eCDF plots**

588 A custom perl script (annotateWithSeeds.pl) was written to identify exact seed matches (reverse complement)
589 to all seeds in all sequences, and to output the total counts of the different types of seeds (generally toxic vs.
590 nontoxic) in the sequences. To compare the presence of toxic and nontoxic seed matches in expression
591 matched survival and non survival genes, a custom R script (makeECDFplot.180615.R) was written that takes
592 as input two different sets of genes (SGs and nonSGs) and the list of the counts of toxic and nontoxic seeds
593 (reverse complement) in all genes, and plots the cumulative distribution function for the count statistics in
594 each gene set. In **Fig. 3b** the ratio of the seed match counts to the 20 most and least toxic seeds in the 5'UTR,
595 CDS, first 1000bp of 3'UTR, and full 3'UTR (not shown) were compared between pairs of 938 expression
596 matched survival and non-survival genes. In **Fig. S5a**, this analysis was repeated with the 100 most and least
597 toxic seeds to both human cell lines. The custom scripts and the input data are available in the cloud-based
598 computational reproducibility platform Code Ocean at https://codeocean.com/capsule/b755ec2b-00d8-4281-
599 9fa1-2a484fd7521b/code. To determine the dependence of mRNAs regulation on miR-34a-5p seed presence
600 in their 3'UTR, a custom R script (makeECDFplot.cetoData.R) was written that takes as input a list of gene
601 sets and a table of logFC expression for those genes upon miR-34a-5p or si34a-5p^{Seed} over-expression. This
602 Rscript then plots the cumulative distribution function for the logFC expression data in each gene set. The

503 custom scripts and the input data are available in the Code Ocean at
504 https://codeocean.com/capsule/31ec8deb-8282-4a90-98e6-b80a0ba881cb/code.
505
506 **Relation between miRNA seed conservation, age and toxicity**
507 Information on miRNA seed family conservation and seed sequence were downloaded at
508 http://www.targetscan.org/vert_71/ from TargetScan Human 7.1. The toxicity of each mature human miRNA
509 arm sequence in the TargetScan dataset was assigned according to the average toxicity induced by the siRNA
510 in HeyA8 and H460 siRNA screens harboring the identical 6mer seed sequence. A list of miRNA ages
511 corresponding to ~1025 miRNA loci was acquired from [31] and was calculated using a modified version of
512 ProteinHistorian [31]. This list was used to assign ages to roughly ~1400 mature miRNA arms found in the
513 TargetScan dataset.
514 TargetScan 7.1 partitions the seed family conservation into four groups: highly conserved (group #2),
515 conserved (group #1), low conservation but still annotated as a miRNA (group #0), and low conservation with
516 the possibility of misannotation (group #-1). Probability density and eCDF plots for the assigned seed-
517 dependent toxicity were generated for each seed family conservation group using ggplot2 in R. Probability
518 density and eCDF plots were also generated to show how young (<10 million years) and old (>800 million
519 years) miRNAs compare in terms of the seed-dependent toxicity. All differences between groups in terms of
520 seed-dependent toxicity (always the average of the toxicity determined in HeyA8 and H460 cells) were
521 analyzed using a two-sample two-tailed Kolmogorov-Smirnov test in R.
522
523 **Assessment of dominant arm seed toxicity**
524 The expression (RPM) of miRNA 5p and 3p arms across 135 tissue samples were collected from MIRMINE
525 [56]. A miRNA was considered expressed if the sum of the normalized reads for both arms was above 5 for each
526 sample. A value of 0 was replaced with 0.01 to avoid a division by 0 error. A miRNA arm was considered the
527 dominant species if its expression was at least 25% greater than the other arm per sample. The dominant arm
528 for each miRNA across all samples was calculated by determining which arm was dominantly expressed in
529 more samples (>50% of samples where the miRNA was considered expressed). The miRNAs that have only
530 one annotated arm in miRBase were considered to have only one dominant arm. The seed toxicity values for
531 each arm were extracted from the 4096 siRNA screen data (% average viability for the human HeyA8 and
532 H460 cells) and used to calculate the ratio between the dominant arm's toxicity and the lesser arm's toxicity;
533 miRNAs that only had one expressed arm were not considered in the analyses shown in **Fig. 5d** but are all
534 included on the website: 6merdb.org.
535 To compare the 6mer seed toxicity between up regulated and down regulated Ago bound miRNA
536 populations in HCT116 cells after Doxo treatment we analyzed the reads obtained from RNA Seq analysis of

537 Ago bound RNAs. After removing the reads that were either shorter than 19 nt or longer than 26 nt in
538 length, the reads were blasted against a miRNA database consisting of all human miRNA mature sequence
539 information obtained from miRBase. A threshold of 100% identity for an at least 16 nt-long stretch without
540 any gaps was set for the BLAST analysis. After discarding sequences with no significant BLAST result, the
541 remaining sequences were trimmed from the 3'end so that all reads were now 19nt in length. This was done to
542 determine for each miRNA the relevant 5' end to obtain the 6mer seed sequence (position 2-7). All the reads
543 that shared the same 5' sequence and miRNA names were collapsed while adding up the number of reads of
544 each sequence in each condition. To compare the 6mer seed toxicity between up and down regulated miRNAs,
545 we calculated the average 6mer seed toxicity for miRNA sequences that were either 1.5 fold up or 1.5 fold
546 down regulated in Doxo treated wt samples compared to medium treated control samples (after removing
547 sequences that had less than 100 collapsed reads in Doxo treated wt samples). In each group miRNAs were
548 ranked according to highest base mean expression and groups were compared. Statistically significance was
549 determined using the Mann-Whitney U test. The comparison was repeated for Drosha k.o. cells where the 5'
550 shifted form of miR-320a-3p was the only miRNA found to be upregulated (1.5 fold).
551

552 **Analysis of miRtron and non-miRtron groups**

553 MiRtrons and non-miRtrons were recently reported [57] and consisted of miRNAs that are listed in miRBase v21
554 as expressing both arms. Comparing 6mer toxicity of miRtrons and non-miRtrons from this list was done as
555 described above for young/old non-conserved/conserved miRNAs. Both arms were considered. Probability
556 density (**Fig. 5c**) and eCDF plots (**Fig. S7c**) were generated to show how miRtrons and non miRtrons compare
557 in terms of the seed-dependent toxicity.
558

559 **Comparing the seed viability of shRNAs derived from the CD95L sequence**

560 An RNAi lethality screen composed of every shRNA sequence that can be derived from the CD95L CDS was
561 conducted previously (*20*). In this screen, toxicity of each shRNA was assessed in two ways: 1) fold
562 underrepresentation of the shRNA after infection with the shRNA-expressing lentivirus compared to its
563 representation in the plasmid pool and 2) fold representation of the shRNA after infection and treatment with
564 doxycycline compared to cells that were infected but did not receive doxycycline. The first analysis allowed
565 us to quantify toxicity associated with "leaky" shRNA expression. The second analysis quantified toxicity
566 associated with strong shRNA expression following treatment with doxycycline.

567 For each shRNA, the average fold downregulation was calculated from both of these toxicity assessments.
568 Then, the seed sequence of each shRNA was extracted and assigned an average viability score, which was a
569 composite of the % viabilities determined in the 4096 siRNA arrayed screen for both HeyA8 and H460 cells.

Pearson's correlation was determined for each CD95L-derived shRNA between its associated fold downregulation in the screen (*20*) and the average seed sequence viability determined in the 4096 siRNA screen.

In addition, the CD95L shRNAs were split into two groups: 1) 137 shRNAs with an average fold downregulation (as determined in the shRNA lethality screen) above 5 and 2) a control group with a matching number of shRNAs whose fold deregulation had an absolute value closest to 0. The average seed viability (as determined from the siRNA screen) was extracted for the shRNAs in these two groups and compared using the two-sample, two-tailed Rank Sum test.

Statistical analyses

Continuous data were summarized as means and standard deviations (except for all IncuCyte experiments where standard errors are shown) and dichotomous data as proportions. Continuous data were compared using t-tests for two independent groups and one-way ANOVA for 3 or more groups. For evaluation of continuous outcomes over time, two-way ANOVA was performed using the Stata 14 software with one factor for the treatment conditions of primary interest and a second factor for time treated as a categorical variable to allow for non-linearity. Comparisons of single proportions to hypothesized null values were evaluated using binomial tests. Statistical tests of two independent proportions were used to compare dichotomous observations across groups. Pearson correlation coefficients (r) and p - values as well as Mann Whitney U test were calculated using StatPlus (v. 6.3.0.5). Kolmogorov–Smirnov two-sample two-sided test was used to compare different probability distributions shown in all density plots and eCDF plots. Wilcoxon rank sum test was used to test for statistical significance in the analysis of the toxicity all miRNA arms in **Fig. 5d**.

The effects of treatment vs. control over time were compared for Drosha k.o. and wild-type cells by fitting regression models that included linear and quadratic terms for value over time, main effects for treatment and cell type, and two- and three-way interactions for treatment, cell-type and time. The three-way interaction on the polynomial terms with treatment and cell type was evaluated for statistical significance since this represents the difference in treatment effects over the course of the experiment for the varying cell types.

Data availability

RNA sequencing data generated for this study is available in the GEO repository: GSE111379 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111379, reviewer access token: srileyyktjctduz) and GSE111363 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111363, reviewer access token: udedukauvjexbmp). All 6mer seed toxicity data of the 4096 siRNA screen in HeyA8, H460, M565, and 3LL cells are available in searchable form at https://6merdb.org. All source data are available upon request.

ACKNOWLEDGEMENTS

We are indebted to Dr. Leon Platanias for his generous support of the seed screen and to Denise Scholtens for help with biostatistics. M.H. was supported by the Intramural Research Program of NIAMS and A.A.S. by the Swedish Research Council postdoctoral fellowship. This work was funded by training grant T32CA009560 (to W.P.), R35CA197450 (to M.E.P.), R50CA221848 (to E.T.B.), and RO1 211916 (to EL).

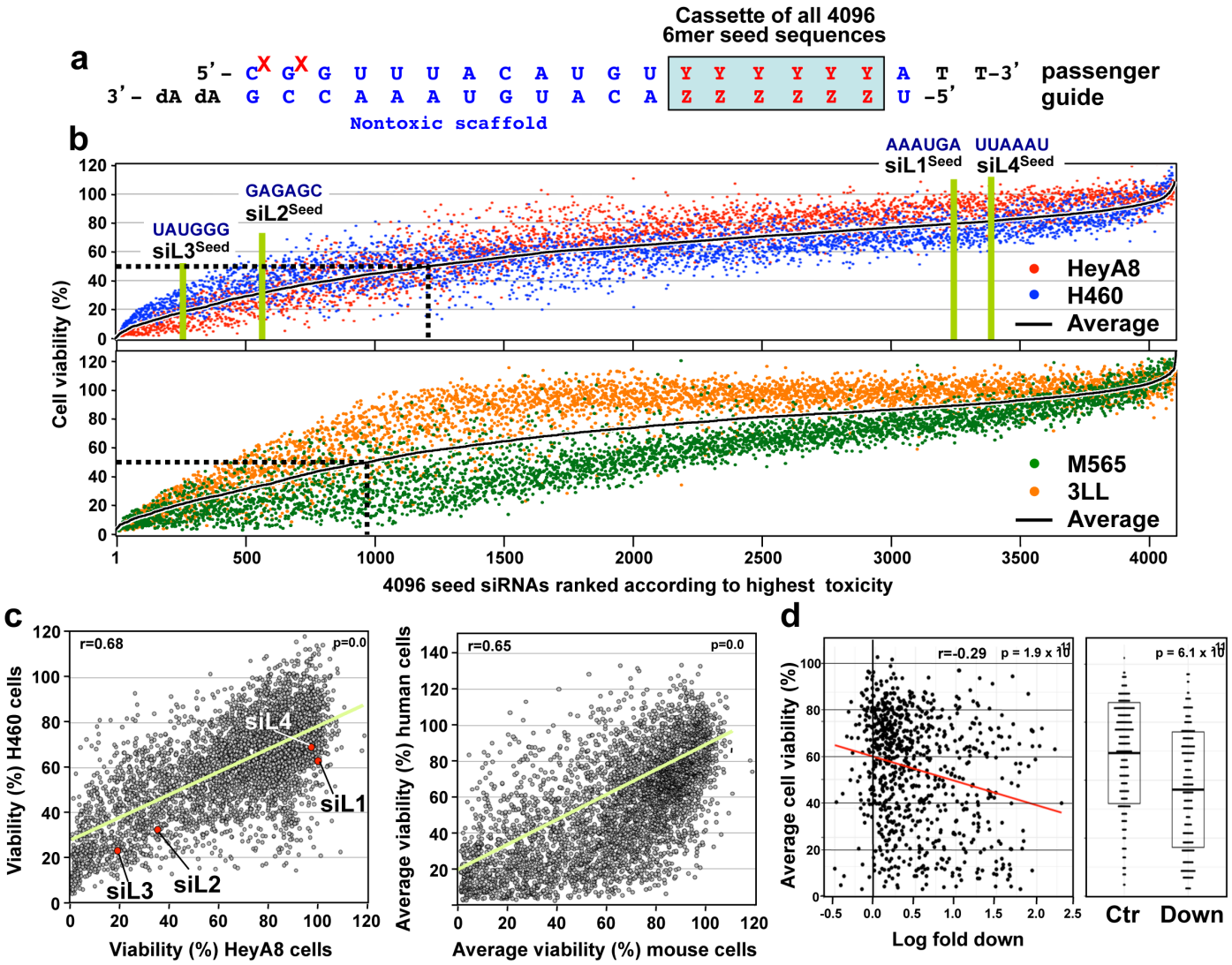
AUTHOR CONTRIBUTIONS

M.E.P., Q.Q.G., W.E.P. and A.E.M. conceived the study. Q.Q.G., W.E.P. performed the majority of the experiments and data analysis. S.C. performed the 6mer seed screens. A.A.S. performed the Ago pull-down experiments. E.T.B. performed data analyses. J.M.P designed and implemented the website. M.H. provided critical conceptual input. M.E.P. designed the study, guided the interpretation of the results and drafted the manuscript. All authors discussed the results, edited and approved the draft and final versions of the manuscript.

DECLARATION OF INTEREST

The authors declare no competing interests.

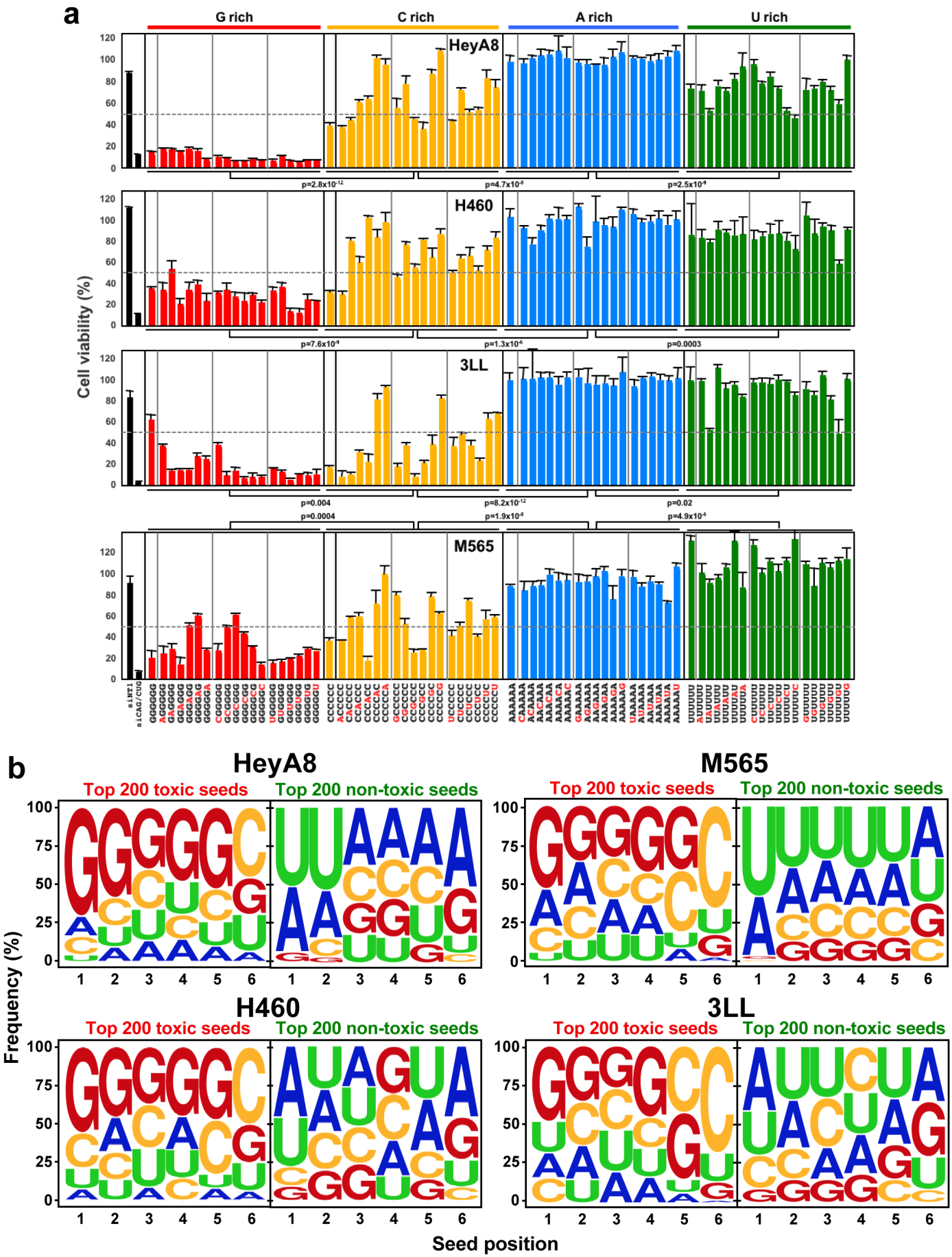
723 **Figures**



724

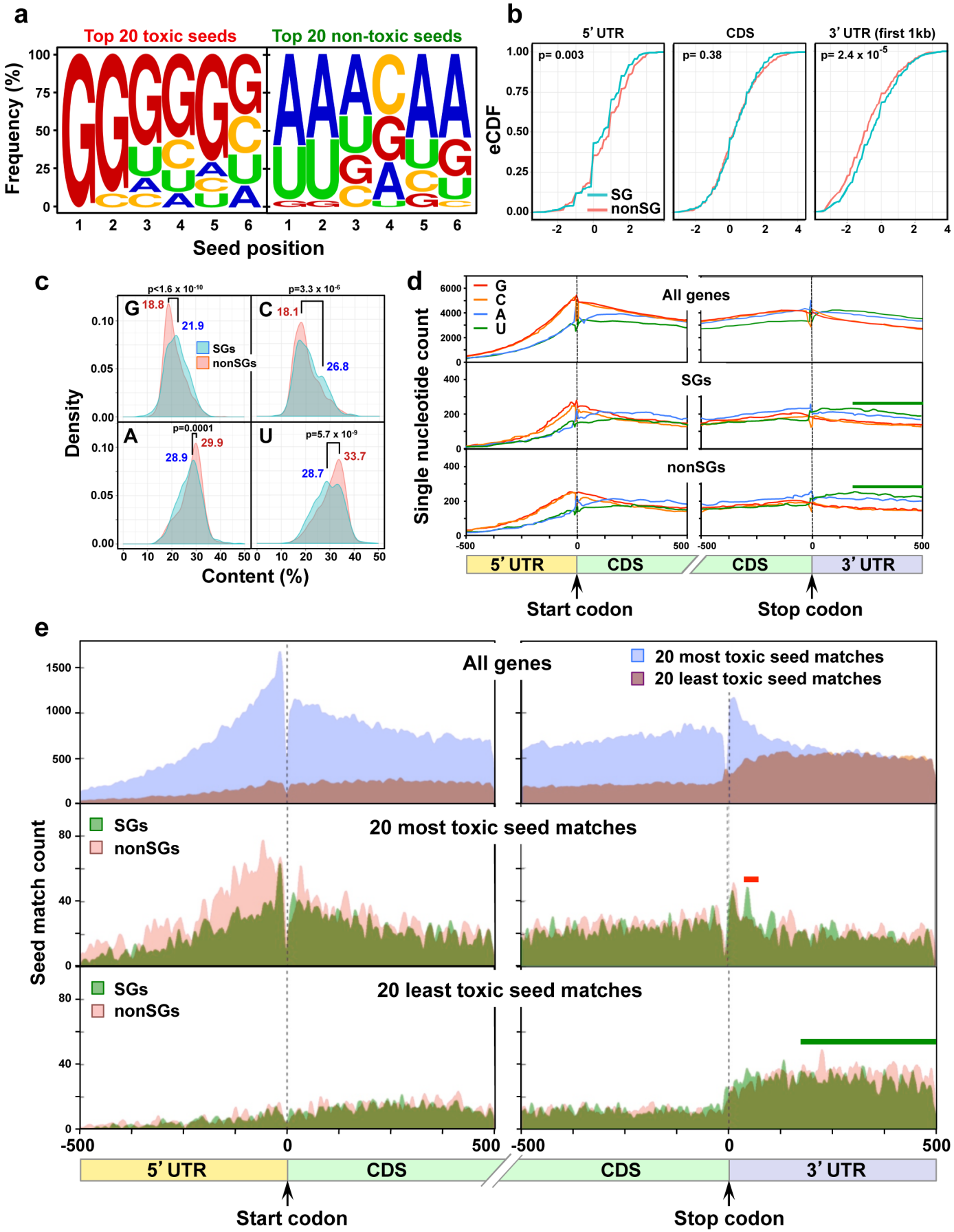
725 **Fig. 1** A comprehensive screen identifies the most toxic 6mer seeds. **a** Schematic of the siRNA backbone
726 used in the 4096 seed duplexes toxicity screen. 2'-O-methylation modifications at position 1 and 2 of the
727 passenger strand were added to prevent passenger strand loading (marked by red Xs) [22]. The nucleotides that
728 remained constant in all the duplexes are in blue. The variable 6mer seed sequence is in red (Y and
729 complementary Z) and boxed in blue. **b** *Top:* Results of the 4096 6mer seed duplex screen in human HeyA8
730 and H460 cells. *Bottom:* Results of the 4096 6mer seed duplex screen in mouse M565 and 3LL cells. Cells
731 were reverse transfected in triplicates in 384 well plates with 10 nM of individual siRNAs. The cell viability
732 of each 6mer seed duplex was determined by quantifying cellular ATP content 96 hours after transfection. All
733 4096 6mer seeds are ranked by the average effect on cell viability of the two cell lines from the most toxic
734 (left) to the least toxic (right). Rankings of the 6mer seeds of four previously characterized CD95L derived
735 siRNAs (siL1, siL2, siL3, and siL4) are highlighted in green. We consider an siRNA highly toxic if it reduces
736 cell viability 90% or more and moderately toxic if it reduces cell viability 50% or more (black stippled line). **c**

737 Regression analysis showing the correlation between the 6mer seed toxicity observed in the human lung
738 cancer cell line H460 (y axis) and the matching 6mer toxicity observed in the human ovarian cancer cell line
739 HeyA8 (x axis) (left) and average toxicity in the two human cell lines (y axis) and two mouse cell lines (x
740 axis) (right). p-values were calculated using Pearson correlation analysis. Toxicity of siRNAs and miRNAs in
741 all four cell lines can be interrogated at https://.6merdb.org. **d** *Left:* Correlation between the $\log_{10}$ (fold down
742 underrepresentation) of all possible shRNAs that can be derived from the mRNA sequence of CD95L
743 following their expression from a DOX-inducible lentiviral vector [20] and the toxicity of their seed sequences
744 as determined in a 4096 arrayed siRNA screen (average of both human cell lines) representing every seed
745 6mer permutation [20]. *Right*: Difference in average seed toxicity between the 137 CD95L-derived shRNAs
746 downregulated at least 5 fold (= toxic) in this screen compared to a size matched group of 137 shRNAs that
747 were the least altered in their abundance in that screen. Pearson correlation coefficient is given as well as p-
748 value (left) and p-value in analysis on the right was calculated using unpaired two arm ttest.
749


a
G rich
C rich
A rich
U rich
HeyA8
H460
3LL
M565
Cell viability (%)
p=2.8x10-12
p=4.7x10-5
p=2.5x10-9
p=7.6x10-9
p=1.3x10-5
p=0.0003
p=0.004
p=8.2x10-12
p=0.02
p=0.0004
p=1.9x10-5
p=4.9x10-5
siNT1
siCAG/CUG
b
HeyA8
M565
H460
3LL
Top 200 toxic seeds
Top 200 non-toxic seeds
Frequency (%)
Seed position

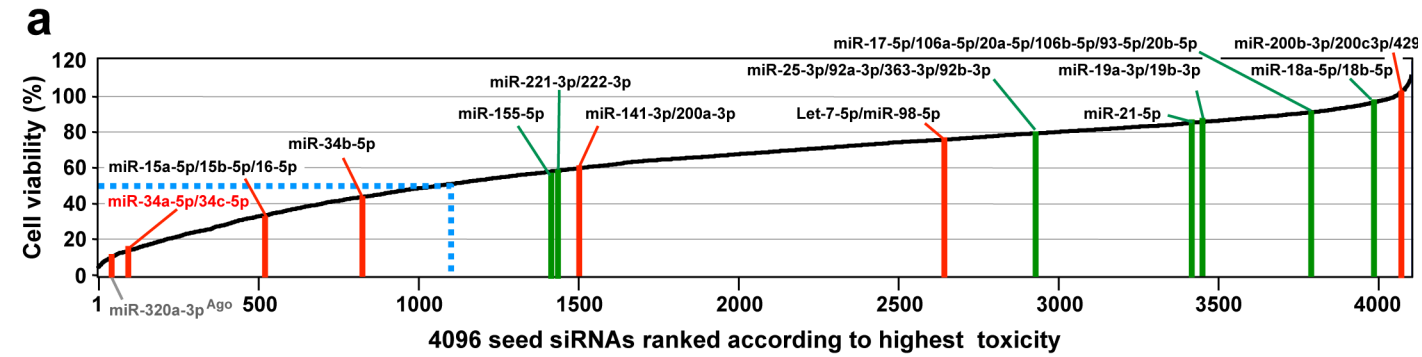
750

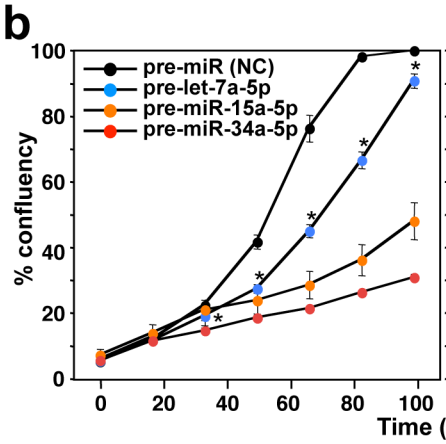
751 **Fig. 2** The most toxic seeds are G rich. **a** Cell viability of the 19 seed duplexes with the highest content
752 (>80%) in the 6mer seed region for each nucleotide in two human and two mouse cell lines. p-values between
753 groups of duplexes were calculated using Student's t-test. **b** Nucleotide composition at each of the 6 seed
754 positions in the top 200 most toxic (left) or the top 200 least toxic (right) seed duplexes in the four cell lines.
755 siRNAs are considered to be toxic when viability is inhibited >50% (grey stippled line).
756

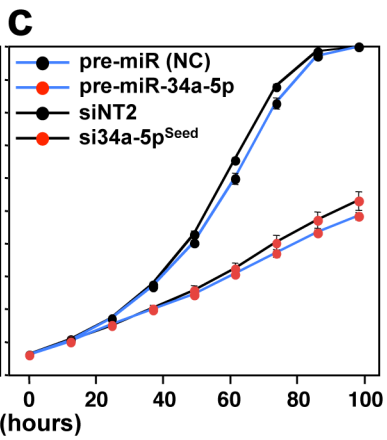

a
Top 20 toxic seeds
Top 20 non-toxic seeds
Frequency (%)
Seed position
b
5' UTR
CDS
3' UTR (first 1kb)
p= 0.003
p= 0.38
p= 2.4 x 10-5
eCDF
SG
nonSG
c
p<1.6 x 10-10
p=3.3 x 10-6
p=0.0001
p=5.7 x 10-9
SGs
nonSGs
Density
Content (%)
d
All genes
SGs
nonSGs
Single nucleotide count
5' UTR
CDS
CDS
3' UTR
Start codon
Stop codon
e
All genes
20 most toxic seed matches
20 least toxic seed matches
SGs
nonSGs
Seed match count
5' UTR
CDS
CDS
3' UTR
Start codon
Stop codon

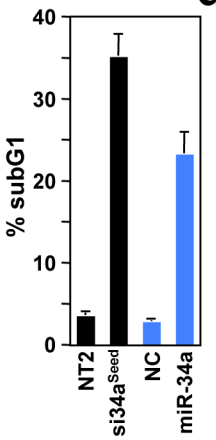
757

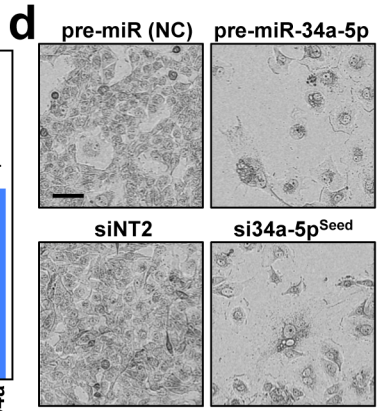
Fig. 3 The most toxic G-rich seed containing duplexes preferentially target housekeeping genes enriched in Cs close to the 3‘UTR start site. **a** Nucleotide composition of 20 seeds that are most and least toxic in both human cell lines (see **Fig. 1b**). **b** eCDF comparing the ratio of occurrences of the 20 most and least toxic 6mer seed matches in the mRNA elements of two sets of expression matched survival genes and nonsurvival genes. Significance was calculated using a two-sample two-sided Kolmogorov–Smirnov (K-S) test. Significance was higher in the analysis shown with the first 1000 bases of each 3'UTR than with the entire 3'UTR (data not shown). **c** Probability density plots comparing the nucleotide content between the groups of expression matched SGs and nonSGs. p-values were calculated using a two-sample two-sided K-S test comparing the density distribution of SGs and nonSGs. Relevant peak maxima are given. **d** Single nucleotide frequency distribution in human mRNAs around the boundaries of the 5'UTR and the beginning of the CDS and the end of the CDS and the beginning of the 3'UTR (shown are 500 bases in each direction). Data are shown for either all human coding genes (top), a set of 938 SGs or an expression matched set of 938 nonSGs. Green horizontal bars, area of reduced A/U content in SGs. **e** Distribution of the seed matches to the 20 most and least toxic 6mer seeds to human cells in human mRNAs around the boundaries of the 5'UTR and the beginning of the CDS and the end of the CDS and the beginning of the 3'UTR (shown are 500 bases in each direction). Data are shown for either all genes (top) or the expression matched SGs and nonSGs (center and bottom). Red horizontal bar, area of enriched toxic seed matches in SGs compared to nonSGs. Green horizontal bar, area of fewer toxic seed matches in SGs.

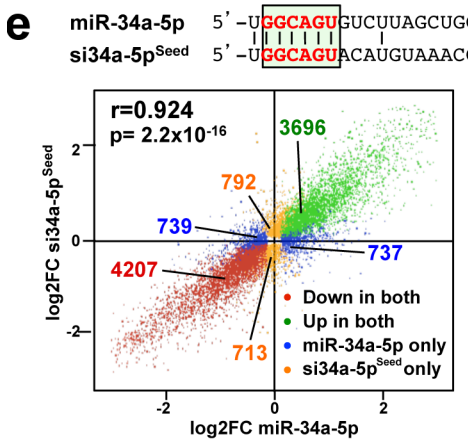


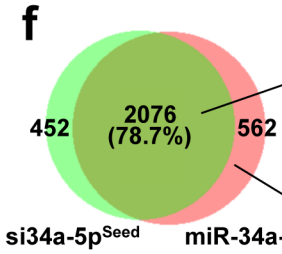






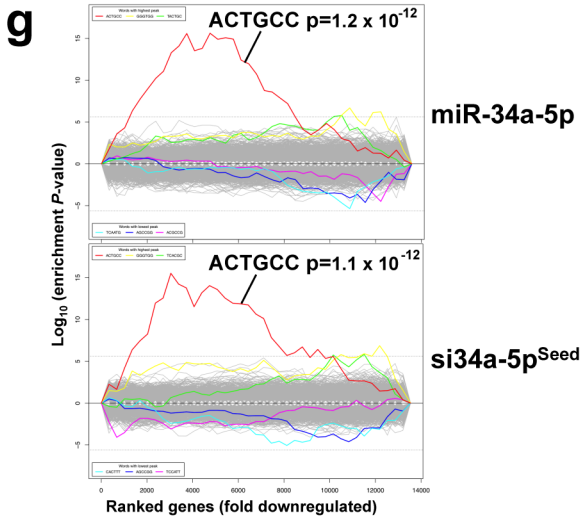


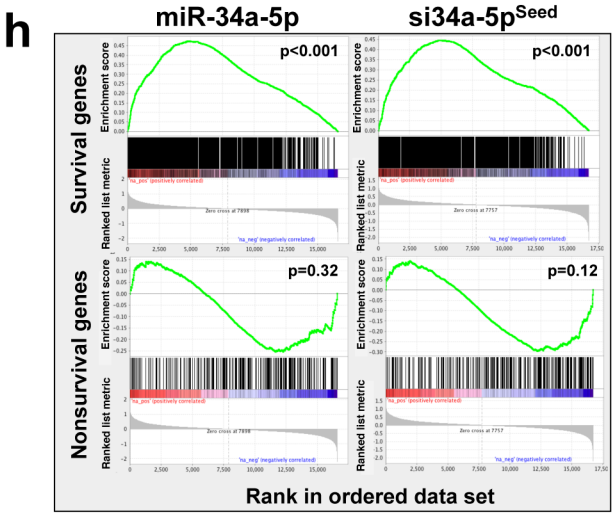




GO Term	Description	P-value	FDR q-value	Enrichment
GO:0022402	cell cycle process	3.25E-27	2.60E-23	2.0
GO:0065004	protein-DNA complex assembly	2.80E-23	1.12E-19	3.8
GO:0071824	protein-DNA complex subunit organization	6.06E-22	1.61E-18	3.6
GO:0006334	nucleosome assembly	3.38E-21	6.75E-18	3.9
GO:1903047	mitotic cell cycle process	8.37E-21	1.34E-17	2.0
GO:0034728	nucleosome organization	4.22E-20	5.62E-17	3.7
GO:0051276	chromosome organization	6.66E-19	7.59E-16	2.7
GO:0006325	chromatin organization	5.75E-16	5.74E-13	4.7
GO:0006342	chromatin silencing	5.37E-15	4.76E-12	4.3
GO:0045814	negative regulation of gene expression, epigenetic	1.16E-14	9.28E-12	4.1
GO:0016458	gene silencing	5.56E-13	4.03E-10	6.9
GO:0051301	cell division	7.46E-13	4.96E-10	2.0
GO:0038111	interleukin-7-mediated signaling pathway	1.64E-12	1.01E-09	19.3
GO:0000183	chromatin silencing at rDNA	3.29E-12	1.88E-09	5.3
GO:0006259	DNA metabolic process	4.15E-12	2.21E-09	1.5

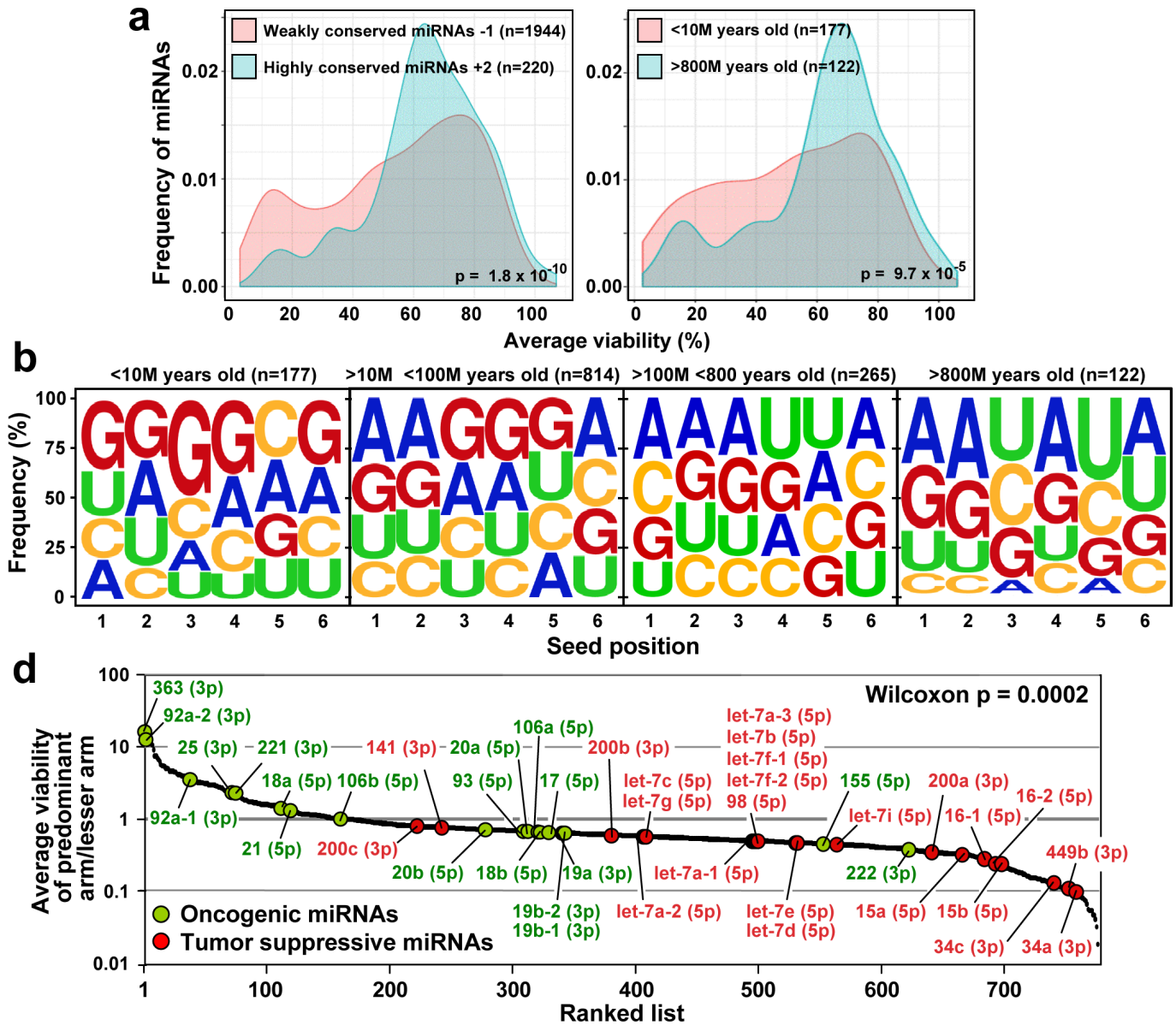
GO Term	Description	P-value	FDR q-value	Enrichment
GO:0015718	monocarboxylic acid transport	4.59E-06	1.82E-02	22.5
GO:0006869	lipid transport	2.57E-05	5.09E-02	14.1
GO:0015850	organic hydroxy compound transport	1.02E-04	1.35E-01	22.5
GO:0015849	organic acid transport	1.83E-04	1.81E-01	6.0
GO:0046942	carboxylic acid transport	1.83E-04	1.45E-01	6.0
GO:0006820	anion transport	4.80E-04	3.17E-01	9.4





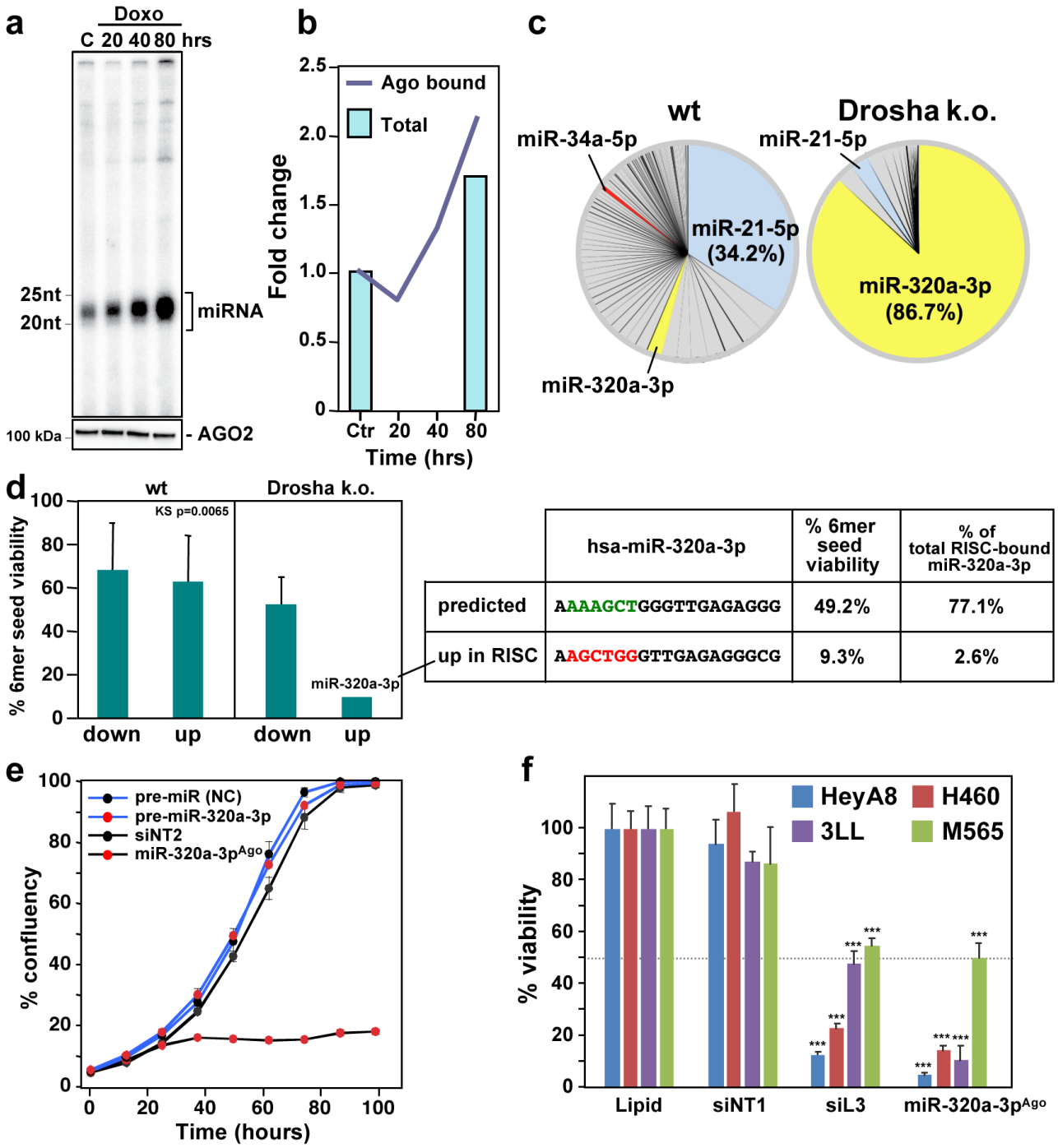
777

Fig. 4 Tumor suppressive miRNAs inhibit cancer cell growth via toxic 6mer seeds. **a** All 4096 6mer seeds ranked from the lowest average viability (highest toxicity) to the highest viability (lowest toxicity) in HeyA8 and H460 cells. Locations of 6mer seeds present in major tumor suppressive (red) or tumor promoting (green) miRNAs are highlighted as individual bars. miRNAs are considered to be toxic when viability is inhibited >50% (blue stippled line). **b** Percent cell confluence over time of HeyA8 cells transfected with 5 nM of either the indicated tumor suppressive miRNA precursors or a miRNA precursor nontargeting control. Data are representative of two independent experiments. Each data point represents mean ± SE of four replicates. *Two-way ANOVA p-value between cells treated with pre-miR-(NC) and pre-let-7a-5p is 0.0000. **c** *Left*: Percent cell confluence over time of HeyA8 parental cells transfected with either pre-miR-34a-5p or si34a-5p^{Seed} and compared to their respective controls (pre-miR (NC) for pre-miR-34a and siNT2 for si34a^{Seed}) at 10 nM. Data are representative of two independent experiments. Each data point represents mean ± SE of four replicates. *Right*: % cell death of the same cells harvested 4 days after transfection. Data are representative of two experiments. Each data point represents mean ± SD of three replicates. **d** Morphology of HeyA8 cells transfected with either pre-miR-34a-5p or si34a-5p^{Seed} compared to their respective controls at 10 nM three days after transfection. **e** *Top*: Alignment of the sequences of miR-34a-5p and miR-34a-5p^{Seed} with the 6mer highlighted. *Bottom:* Comparison of deregulated mRNAs (adjusted $p<0.05$, RPM > 1) in HeyA8 cells 48 hrs after transfection with either miR-34a-5p or si34a-5p^{Seed}. Pearson correlation p-value is given. **f** Overlap of RNAs detected by RNA-Seq downregulated in HeyA8 cells (>1.5 fold) 48 hrs after transfection with either si34a-5p^{Seed} or miR-34a-5p when compared to either siNT2 or a nontargeting pre-miR, respectively. *Right*: Results of a GOrilla gene ontology analyses of the genes downregulated in both cells transfected with miR-34a-5p or si34a-5p^{Seed} (top, significance of enrichment $<10^{-11}$), or only in cells transfected with miR-34a-5p (bottom, significance of enrichment $<10^{-4}$). **g** Sylamer plots for the list of 3'UTRs of mRNAs in cells treated with either miR-34a-5p (top) or si34-5p^{Seed} (bottom) sorted from down-regulated to up-regulated. The most highly enriched sequence is shown which, in each case, is the 6mer seed match of the introduced 6mer seed. Bonferroni-adjusted p-values are shown. **h** Gene set enrichment analysis for a group of 1846 survival genes (top 4 panels) and 416 non-survival genes (bottom 2 panels) [20] after transfecting HeyA8 cells with either miR-34a-5p or si34a-5p^{Seed}. siNT1 and a nontargeting premiR served as controls, respectively. p-values indicate the significance of enrichment.


a
Weakly conserved miRNAs -1 (n=1944)
Highly conserved miRNAs +2 (n=220)
p = 1.8 x 10-10
<10M years old (n=177)
>800M years old (n=122)
p = 9.7 x 10-5
Frequency of miRNAs
Average viability (%)
b
<10M years old (n=177)
>10M <100M years old (n=814)
>100M <800 years old (n=265)
>800M years old (n=122)
Frequency (%)
Seed position
d
Wilcoxon p = 0.0002
363 (3p)
92a-2 (3p)
25 (3p)
221 (3p)
141 (3p)
20a (5p)
106a (5p)
200b (3p)
let-7a-3 (5p)
let-7b (5p)
let-7f-1 (5p)
let-7f-2 (5p)
98 (5p)
155 (5p)
200a (3p)
16-2 (5p)
16-1 (5p)
449b (3p)
18a (5p)
106b (5p)
93 (5p)
17 (5p)
let-7c (5p)
let-7g (5p)
let-7i (5p)
92a-1 (3p)
21 (5p)
200c (3p)
20b (5p)
18b (5p)
19a (3p)
let-7a-1 (5p)
222 (3p)
19b-2 (3p)
19b-1 (3p)
let-7a-2 (5p)
let-7e (5p)
let-7d (5p)
15a (5p)
15b (5p)
34c (3p)
34a (3p)
Oncogenic miRNAs
Tumor suppressive miRNAs
Average viability of predominant arm/lesser arm
Ranked list
c
miRtrons (n=428)
nonmiRtrons (n=1310)
p = 2.2 x 10-16
Frequency of miRNAs
Average viability (%)
e
-5p
-3p
miR-363
2.7%
77.3%
miR-92a-2
3.5%
77.3%
miR-25
18.6%
77.3%
miR-34a
12.8%
71.4%
miR-34c
12.8%
53.2%
miR-449b
12.8%
64.6%

307

Fig. 5 Toxic 6mer seeds and the evolution of cancer regulating miRNAs. **a** Probability density plot of cell viability of the 6mer seeds of either highly conserved (from humans to zebrafish) or poorly conserved miRNA seed families (left panel, total number of mature miRNAs = 2164) or of very old (>800 Million years) miRNAs or very young (<10 Million years) miRNAs (right panel, total number of miRNAs = 299). For the analysis on the right, miRNA arms with identical sequences (gene duplications) were collapsed and counted as one arm. Two-sample two-sided K-S test was used to calculate p-values. **b** Change in nucleotide composition in the 6mer seeds of miRNAs of different ages. **c** Probability density plot of cell viability of the 6mer seeds of mature miRtrons or non-miRtrons. miRNAs with identical sequences (gene duplications) were collapsed and counted as one seed. Two-sample two-sided K-S test was used to calculate p-value. **d** 780 miRNAs (**Supplementary Table 4**) ranked according to the ratio of viability of the seed (as determined in the seed screen) of the guide strand and the lesser-expressed arm. Established oncogenic miRNAs are shown in green, tumor-suppressive miRNAs are shown in red. The guide strand is given for each miRNA (in parenthesis). p-value of the distribution of oncogenic versus tumor suppressive miRNAs was calculated using two-sample two-sided K-S test. **e** Cumulative read numbers from the 5p or the 3p arm (according to miRBase.org) of three oncogenic and three tumor-suppressive miRNAs with the highest (top three) or lowest (bottom three) ratio of the viability of the guide strand versus the lesser arm. The viability numbers of the matching 6mer seeds according to the siRNA 6mer seed screen are given. The sequences of the mature 5p or 3p arms are boxed in blue and black, respectively. Toxic seeds are shown in red, non-toxic ones in green.



	hsa-miR-320a-3p	% 6mer seed viability	% of total RISC-bound miR-320a-3p
predicted	AAAAGCTGGGTTGAGAGGG	49.2%	77.1%
up in RISC	AAGCTGGGTTGAGAGGGCG	9.3%	2.6%

Fig. 6 Genotoxic drugs cause upregulation of tumor suppressive miRNAs. **a** *Top*: Plot of radiolabeled RNAs pulled down with the Ago proteins from HeyA8 cells treated with doxorubicin (Doxo) for 20, 40, or 80 hours. *Bottom*: Western blot for the pulled down AGO2 of the same samples shown above. The images are representative of two biological duplicates. **b** Fold change of the total reads of Ago bound small RNAs after 20, 40, or 80 hours of Doxo treatment compared to the control sample from Ago-IP sequencing data (Ago bound). Fold change of the total reads of cytosolic small RNAs in HeyA8 cells treated with Doxo for 80 hours compared to the control sample from small RNA-Seq data is given (Total). Data are the combination of

biological duplicates. **c** Pie charts showing the composition of miRNAs bound to Ago proteins after 50 hrs Doxo treatment in HCT116 wild-type (left) or Drosha k.o. cells (right). **d** *Left*, 6mer seed viability (average between HeyA8 and H460 cells) of the Ago-bound miRNAs most up and downregulated in wt or Drosha k.o. cells after Doxo treatment. *Right*, Comparison of the predicted (and most abundant) sequence of miR-320a-3p and Ago-bound sequence of miR-320a-3p and their average viability found most upregulated in Drosha k.o. cells after Doxo treatment. Shown is variance of two biological replicates. Mann-Whiney U test was used to calculate p-value. **e** Percent cell confluence over time of HeyA8 cells transfected with 5 nM of controls, pre-miR-320a or an siRNA duplex that corresponds to the shifted form of miR-320a-3p (si-miR-320a-3p^{Ago}) that was found to be upregulated and bound to Ago proteins upon Doxo treatment. Data are representative of two independent experiments. Each data point represents mean ± SE of four replicates. **f** Viability changes (ATP content) in four cell lines 48 hrs after transfection with Lipid only, 10 nM of siNT1, siL3, a nontargeting pre-miR, or miR-320a-3p^{Ago} - the only shared upregulated miRNA in HeyA8 cells, HCT116 wild-type, and HCT116 Drosha k.o. cells - after Doxo treatment. p-values were determined using Student's t-test. *** $p<0.0001$. Samples were performed in triplicate (siNT1, siL3), 6 repeats (miR-320a-3p^{Ago}) and 8 repeats (Lipid).

REFERENCES

1. Lee Y, *et al*. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060 (2004).

2. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-3027 (2004).

3. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-3016 (2003).

4. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366 (2001).

5. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-838 (2001).

6. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209-213 (2008).

7. Leuschner PJ, Ameres SL, Kueng S, Martinez J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep* **7**, 314-320 (2006).

8. Schirle NT, MacRae IJ. The crystal structure of human Argonaute2. *Science* **336**, 1037-1040 (2012).

9. Eulalio A, Huntzinger E, Izaurralde E. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* **15**, 346-353 (2008).

10. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell* **115**, 787-798 (2003).

11. Lai EC. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**, 363-364 (2002).

12. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).

13. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).

14. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-269 (2006).

15. Balatti V, Pekarky Y, Rizzotto L, Croce CM. miR deregulation in CLL. *Adv Exp Med Biol* **792**, 309-325 (2013).

16. Slabakova E, Culig Z, Remsik J, Soucek K. Alternative mechanisms of miR-34a regulation in cancer. *Cell Death Dis* **8**, e3100 (2017).

17. Hua YJ, Larsen N, Kalyana-Sundaram S, Kjems J, Chinnaiyan AM, Peter ME. miRConnect 2.0: Identification of antagonistic, oncogenic miRNA families in three human cancers. *BMC Genomics* **14**, 179 (2013).

18. Concepcion CP, Bonetti C, Ventura A. The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer journal* **18**, 262-267 (2012).

19. He X, He L, Hannon GJ. The guardian's little helper: microRNAs in the p53 tumor suppressor network. *Cancer Res* **67**, 11099-11101 (2007).

20. Putzbach W, *et al*. Many si/shRNAs can kill cancer cells by targeting multiple survival genes through an off-target mechanism. *eLife* **6**, e29702 (2017).

21. Murmann AE, *et al*. Induction of DISE in ovarian cancer cells in vivo. *Oncotarget* **8**, 84643-84658 (2017).

22. Murmann AE, *et al*. Small interfering RNAs based on huntingtin trinucleotide repeats are highly toxic to cancer cells. *EMBO Rep* **19**, e45336 (2018).

23. Putzbach W, *et al*. CD95L mRNA is toxic to cells *bioRxive*, https://doi.org/10.1101/330324 (2018).

24. Hadji A, *et al*. Death induced by CD95 or CD95 ligand elimination. *Cell Reports* **10**, 208-222 (2014).

25. Bramsen JB, *et al*. A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res* **37**, 2867-2881 (2009).

26. Peter ME. Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle* **8**, 843-852 (2009).

27. Agostini M, Knight RA. miR-34: from bench to bedside. *Oncotarget* **5**, 872-881 (2014).

28. Patel M, Peter ME. Identification of DISE-inducing shRNAs by monitoring cellular responses. *Cell Cycle* **17**, 506-514 (2018).

29. van Dongen S, Abreu-Goodger C, Enright AJ. Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods* **5**, 1023-1025 (2008).

30. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).

31. Patel VD, Capra JA. Ancient human miRNAs are more likely to have broad functions and disease associations than young miRNAs. *BMC Genomics* **18**, 672 (2017).

32. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. *Mol Cell* **28**, 328-336 (2007).

33. Meijer HA, Smith EM, Bushell M. Regulation of miRNA strand selection: follow the leader? *Biochem Soc Trans* **42**, 1135-1140 (2014).

34. Concepcion CP, *et al*. Intact p53-dependent responses in miR-34-deficient mice. *PLoS Genet* **8**, e1002797 (2012).

35. Meredith AM, Dass CR. Increasing role of the cancer chemotherapeutic doxorubicin in cellular metabolism. *J Pharm Pharmacol* **68**, 729-741 (2016).

36. Yeung TK, Germond C, Chen X, Wang Z. The mode of action of taxol: apoptosis at low concentration and necrosis at high concentration. *Biochem Biophys Res Commun* **263**, 398-404 (1999).

37. Yoo SH, *et al*. Etoposide induces a mixed type of programmed cell death and overcomes the resistance conferred by Bcl-2 in Hep3B hepatoma cells. *Int J Oncol* **41**, 1443-1454 (2012).

38. Kwon HK, *et al*. Etoposide Induces Necrosis Through p53-Mediated Antiapoptosis in Human Kidney Proximal Tubule Cells. *Toxicol Sci* **148**, 204-219 (2015).

39. Wang D, Lippard SJ. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov* **4**, 307-320 (2005).

40. Chang BD, *et al*. A senescence-like phenotype distinguishes tumor cells that undergo terminal proliferation arrest after exposure to anticancer agents. *Cancer Res* **59**, 3761-3767 (1999).

41. Eom YW, *et al*. Two distinct modes of cell death induced by doxorubicin: apoptosis and cell death through mitotic catastrophe accompanied by senescence-like phenotype. *Oncogene* **24**, 4765-4777 (2005).

42. Hauptmann J, *et al*. Biochemical isolation of Argonaute protein complexes by Ago-APP. *Proc Natl Acad Sci U S A* **112**, 11841-11845 (2015).

43. Kim YK, Kim B, Kim VN. Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc Natl Acad Sci U S A* **113**, E1881-1889 (2016).

44. Walz AL, *et al*. Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. *Cancer Cell* **27**, 286-297 (2015).

45. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* **39**, 673-677 (2007).

46. Chandradoss SD, Schirle NT, Szczepaniak M, MacRae IJ, Joo C. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **162**, 96-107 (2015).

47. Di Martino MT, *et al*. Synthetic miR-34a mimics as a novel therapeutic agent for multiple myeloma: in vitro and in vivo evidence. *Clin Cancer Res* **18**, 6260-6270 (2012).

48. Beg MS, *et al*. Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Invest New Drugs* **35**, 180-188 (2017).

49. Cao W, *et al*. miR-34a regulates cisplatin-induce gastric cancer cell death by modulating PI3K/AKT/survivin pathway. *Tumour Biol* **35**, 1287-1295 (2014).

50. Navarro F, Lieberman J. miR-34 and p53: New Insights into a Complex Functional Relationship. *PLoS One* **10**, e0132767 (2015).

51. Alfano L, Costa C, Caporaso A, Antonini D, Giordano A, Pentimalli F. HUR protects NONO from degradation by mir320, which is induced by p53 upon UV irradiation. *Oncotarget* **7**, 78127-78139 (2016).

52. Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* **123**, 1133-1146 (2005).

53. Zare H, Khodursky A, Sartorelli V. An evolutionarily biased distribution of miRNA sites toward regulatory genes with high promoter-driven intrinsic transcriptional noise. *BMC Evol Biol* **14**, 74 (2014).

54. Ceppi P, *et al*. CD95 and CD95L promote and protect cancer stem cells. *Nature Commun* **5**, 5238 (2014).

55. Hafner M, Renwick N, Farazi TA, Mihailovic A, Pena JT, Tuschl T. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods* **58**, 164-170 (2012).

56. Panwar B, Omenn GS, Guan Y. miRmine: a database of human miRNA expression profiles. *Bioinformatics* **33**, 1554-1560 (2017).

57. Rorbach G, Unold O, Konopka BM. Distinguishing mirtrons from canonical miRNAs with data exploration and machine learning methods. *Sci Rep* **8**, 7560 (2018).