6mer Seed Toxicity Determines Strand Selection in miRNAs

Quan Q. Gao[1,6], William E. Putzbach[1,6], Andrea E. Murmann[1], Siquan Chen[2], Giovanna Ambrosini[3], Johannes M. Peter[4], Elizabeth T. Bartom[5], 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] School of Life Sciences, EPFL -Swiss Federal Institute of Technology Lausanne, ISREC - Swiss Institute for Experimental Cancer Research, CH-1015 Lausanne, Switzerland, [4] DigiPen Institute of Technology, Redmond, WA 98052

[6] Authors share first authorship

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

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

SUMMARY

Many siRNAs and shRNAs 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 RNA backbone revealed a preference for guanine in positions 1-3 and a GC content of >80% of the 6mer seed in the most toxic siRNAs. These 6mer seed containing siRNAs exert their toxicity by targeting survival genes which contain GC-rich 3'UTRs. The master tumor suppressor miRNA miR-34a was found to be toxic through such a G-rich 6mer seed suggesting that certain tumor suppressive miRNAs use a toxic 6mer seed to kill cancer cells. An analysis of all mature miRNAs suggests that most miRNAs evolved to avoid guanine at the 5' end of the 6mer seed sequence of the predominantly expressed arm. In contrast, for many tumor suppressive miRNAs the predominant arm 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 19-21nt long double stranded RNAs that represses expression of target mRNAs that harbor reverse complementarity to the antisense/guide strand of the small RNA. This results in degradation of the targeted mRNA or in translational repression. RNAi-active guide RNAs can be 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) (Lee et al., 2004). Pri-miRNAs are first processed by the Drosha/DGCR8 Microprocessor complex into pre-microRNAs (Han et al., 2004) which are exported from the nucleus to the cytoplasm by Exportin 5 (Yi, Qin, Macara, & Cullen, 2003). Once in the cytoplasm, Dicer processes them further (Bernstein, Caudy, Hammond, & Hannon, 2001; Hutvagner et al., 2001) and these mature dsRNA duplexes are then loaded into Argonaute (Ago) proteins to form the RNA-induced silencing complex (RISC) (Y. Wang, Sheng, Juranek, Tuschl, & Patel, 2008). The sense/passenger strand is ejected/degraded, while the guide strand remains associated with the RISC (Leuschner, Ameres, Kueng, & Martinez, 2006). While si/shRNAs are usually designed with 100% reverse complementarity to their intended targets and induce AGO2 dependent target degradation (Schirle & MacRae, 2012), cleavage-independent RNAi as exerted by miRNAs causes deadenylation/degradation or translational repression via interaction between AGO2 and TNRC6 family proteins that in turn recruit the CCR4-NOT deadenylase complex (Eulalio, Huntzinger, & Izaurralde, 2008). RNAi 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 the target RNA (Lai, 2002; Lewis, Shih, Jones-Rhoades, Bartel, & Burge, 2003). This seed-based targeting is restricted to binding sites located in the 3'UTR of a target mRNA (Baek et al., 2008; Selbach et al., 2008).

miRNAs play important roles in cancer (Esquela-Kerscher & Slack, 2006). While the function of a miRNA depends on the nature of its targets, some miRNA families have predominantly tumor suppressive or oncogenic activities. The most oncogenic miRNAs include the miR-17~93 family, miR-21, miR-155 and miR-221/222 (Hua et al., 2013). The most consistent tumor suppressors are the miR-34 family (three members), the let-7 family (13 members), and the miR-15/16 family (3 members) (Esquela-Kerscher & Slack, 2006; Hua et al., 2013). miRNAs of the miR-34 family are induced by p53 (He, He, & Hannon, 2007). Hundreds of cell- and animal-based studies support the function of miR-34a as a master regulator of tumor suppression and provide the rationale for miR-34a as a potential therapeutic reagent (Agostini & Knight, 2014).

We recently discovered that many si- and shRNAs can kill all cancer cells through RNAi by targeting the 3'UTR of critical survival genes (Putzbach, Gao, Patel, van Dongen, et al., 2017). This toxicity involves activation of multiple cell death pathways in parallel, with a ROS mediated necrotic form of mitotic catastrophe at its core (Hadji et al., 2014). In addition, cells suffer from extreme stress followed by signs of genomic instability and aneuploidy. Finally, cancer cells have a hard time developing resistance to this treatment both *in vitro* and when treated *in vivo* (Murmann et al., 2017). We reported that a 6mer seed sequence in the toxic siRNAs is sufficient for effective killing (Putzbach, Gao, Patel, van Dongen, et al., 2017).

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. We report that the most toxic seeds are G-rich

with the highest G content towards the 5' end of the seed. We demonstrate that their toxicity can be explained by targeting survival genes with high GC content in their 3'UTR. We also report that many tumor suppressive miRNAs contain G-rich 6mer seeds which are more toxic in our screen than the seeds found in oncogenic miRNAs. miR-34a was found to be one of the most toxic miRNAs and most of its toxicity comes from its 6mer seed sequence. We found that mature miRNAs from older and more conserved miRNAs contain less toxic seeds. Our data provide an explanation as to why certain miRNAs predominantly express the 5p arm whereas others express mostly the 3p arm. 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 contains 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 have evolved to kill cancer cells through a toxic G-rich 6mer seed.

RESULTS

Identifying the Most Toxic 6mer Seeds

To test the effect of any 6mer seed present in the guide strand of an siRNA on the survival of cancer cells we recently designed a neutral 19mer oligonucleotide scaffold with two nucleotide 3' overhangs (Murmann et al., 2018). 6mers could be inserted at positions 2-7 of the guide strand. To block loading of the passenger strand into the RISC complex the sense strand was modified at positions 1 and 2 by 2'-O-methylation (**Figure 1A**). Transfection efficiency and conditions were optimized for each cell line used. To determine the general rules of targeting that underlie the observed toxicity, we screened all possible 4096 6mer seed sequences in this 19mer scaffold for toxicity. This allowed us to rank all 4096 6mer seeds according to their toxicity (**Figure 1B**, **Table S1,** and **6merdb.org**). This activity was likely mediated by RNAi, as knockdown of AGO2 abolished the toxicity of the two most toxic siRNA duplexes (**Figure S1A**). Consistent with the 6mer seed of the siRNAs mediating toxicity, the seeds of 4 previously tested siRNAs derived from CD95L (Hadji et al., 2014) in this screen were about as toxic as the full length siRNAs, with $siL3^{Seed}$ being most toxic followed by $siL2^{Seed}$ and less or no toxicity associated with $siL4^{Seed}$ and $siL1^{Seed}$ (**Figure 1B**). There was substantial agreement in the toxicity of the transfected 4096 seed duplexes between the human HeyA8 cells and the mouse liver cancer cell line M565 (**Figure S1B**), suggesting some of the rules governing 6mer seed toxicity are conserved between mouse and human. We noticed the nucleotide composition of the two seeds of the toxic siRNAs previously discovered (siL3 and siL2) had a higher G content than that of the seeds of the two non-toxic siRNAs (siL4 and siL1, **Figure 1B**). By analyzing the screen results of the 4096 seeds in both the human and the mouse cell line (**Figure S1C**), we found that G-rich seeds appeared to be more toxic than C-rich seeds. The least toxicity was found with seeds with a high A content. To test the effect of high G content on toxicity directly, we selected and retested the 19 seed duplexes with the highest content (>80%) for each of the four nucleotides and tested these 76 duplexes in two human (HeyA8 and H460, lung cancer) and two mouse (M565 and 3LL, lung cancer) cell lines (**Figure 1B**). The reanalysis also allowed us to determine the reproducibility of the results obtained in the large screen (which for technical reasons had to be performed in three sets). Both the data on the HeyA8 and the M565 cells were highly reproducible especially for the most toxic seeds (**Figure S2A**). 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 (**Figure 1B**). 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 activity (Bramsen et al., 2009). An example of the nucleotide representation in each of the 6 seed positions in one of the commercial siRNA libraries is shown in **Figure S2B** (left panel). In contrast, our complete set of 6mer seed duplexes allowed to test the contributions of all four nucleotides in each of the 6 seed positions (**Figure S2B**, right panel). Analysis of the most C-rich seeds revealed that the identity of the nucleotide at different seed positions influences toxicity (purple box in **Figure 1B**). Their toxicity decreased in all four cell lines when a single C to A replacement was moved from the 5' to the 3' end of the seed (stippled grey lines in **Figure 1C**). To determine the nucleotide content of the most toxic seed, we averaged the nucleotide content at each of the 6 positions of either the 200 most or 200 least toxic seed duplexes

for HeyA8 cells (highlighted in red and green in **Figure 1B**) or for M565 cells (**Table S1**). In both cell lines, we noticed that a high G content towards the 5' end of the seed was the most toxic, while C was most toxic in position 6 (**Figure 1D**). In contrast, nontoxic seeds were much more A and U-rich again with an asymmetry of U predominating toward the 5' end and A at the 3' end of the seed.

siRNAs with Toxic Seeds Target Housekeeping Genes Enriched in C-rich Sequences

Seed sequences rich in G towards their 5' end are most toxic, which suggests they target genes important for cell survival carrying seed matches rich in C in their 3' UTR. To determine whether such genes existed, we performed a Matrix Motif search using the PMWScore tool (PMWTools, http://ccg.vital-it.ch/pwmtools). To be most stringent, the nucleotide composition of the 20 most toxic seeds (**Figure 2A**) was used to generate a matrix and all human 3'UTRs were screened for the best single hits (**Table S2**). When this list, ranked according to the highest score, was subjected to a gene ontology analysis, using GOrilla, the GO terms with the lowest p-values ($<10^{-11}$) were consistent with the affected genes being important for cell survival (**Figure 2B**). They included a number of biosynthetic and metabolic processes. This was not found when the ranking of the list was reversed (**Figure S3A**). To determine whether genes with the highest scores carried specific seed matches, would be targeted by G-rich seeds, or were overall GC-rich, we compared the nucleotide content of the 3'UTR of genes with the highest PWM score (>400, 4288 genes) with a group of the lowest scoring genes of a similar size (<90, 4834 genes) (**Figure S3B**). Compared to the average nucleotide content of all 3'UTRs which are known to be rich in A and U (Mignone, Gissi, Liuni, & Pesole, 2002), the genes with a PWM score >400 had a higher G content and an even higher C content when compared to the genes with a score <90. This suggested that the genes targeted by G-rich seed-containing siRNAs did contain motifs with a higher C content but they were also overall somewhat richer in G. This also suggested that survival/housekeeping genes would have a higher GC content than nonsurvival genes. An analysis of a group of ~1800 survival genes and ~400 nonsurvival genes confirmed this prediction and a subpopulation of survival genes had a higher C content (different peak maxima in **Figure 2C**). To determine whether GC-rich sites occurred at specific locations within 3'UTRs, we plotted the position of the first best match in 3'UTRs identified with either the toxic or the nontoxic matrix (**Figure 2D**). Consistent with an analysis of single nucleotide distribution in all human coding genes (Louie, Ott, & Majewski, 2003), we found the first best matches of GC-rich targets complementary to the toxic matrix to be within 100 nts of the stop codon whereas sequence motifs complementary to the nontoxic matrix were further away (>200 nts). When testing the subset of survival and nonsurvival genes, it appeared the first best seed matches within the toxic matrix of the 3'UTRs of survival genes were closer to the 3'UTR start site than in nonsurvival genes. This big difference in peak maxima was not seen when the analysis was repeated with the nontoxic matrix (**Figure S3C**).

We had previously postulated that 6mer seed toxicity could be an anticancer mechanism and predicted that small RNAi active guide RNAs would be involved in killing cancer cells (Putzbach, Gao, Patel, Haluck-Kangas, et al., 2017). Consistent with this hypothesis, an analysis of published data on point mutations in the exomes of 27 different human cancer showed that losing a C is the predominant point mutation across all cancers (Lawrence et al., 2013). Frequencies of such point mutations ranged from >60% (in acute myeloid leukemia) to >90% (in cervical cancer) (**Figure 2E**). Remarkably, the frequency of the loss of C was independent of the mutational load of cancers.

We noticed that two of the five significantly enriched Top Regulator Effect Networks of genes identified in the PWM analysis when analyzed with Ingenuity Pathway Analysis (IPA) were regulated by two miRNAs with G-rich seed sequences (**Figure S4**). They are predicted to repress expression of genes that promote proliferation of tumor cells, DNA and RNA transcription, and cellular homeostasis. The two 6mer seeds of the miRNAs were highly toxic in our screen (**Figure S4, 6merdb.org**). This result raised the question of whether certain miRNAs could be killing cancer cells by targeting survival and housekeeping genes using toxic 6mer seeds.

Tumor Suppressive miR-34a Kills Cancer Cells through Its Toxic 6mer Seed

To test whether certain miRNAs could kill cancer cells through the toxic 6mer seeds we identified, we analyzed the seed toxicity determined in our screen of all known ~2600 mature miRNAs expressed as either a 3p or 5p

version (**Figure 3A**, **6merdb.org**). While none of the 6mer seeds present in the most oncogenic miRNAs (miR-221/222, miR-21, miR-155, the miR-17~92 cluster (miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, and miR-92a), its paralogues the miR-106b~25 cluster (miR-106b, miR-93 and miR-25), and the miR-106a~363 cluster (miR-106a, mR-18b, miR-20b, miR-19b-2, miR-92-2 and miR-363) (Concepcion, Bonetti, & Ventura, 2012)) were toxic (reduced viability >50%, stippled line in **Figure 3A**), two of the major tumor suppressive miRNA families, miR-15/16 and miR-34a/c and miR-34b contained toxic seeds. This suggests these two families were killing cancer cells through toxic 6mer seeds. Interestingly, two other major tumor suppressive families, let-7 and miR-200, were not found to contain toxic G-rich seeds, suggesting that they may be tumor suppressive through other mechanisms such as inducing and maintaining cell differentiation (Peter, 2009).

When transfecting the pre-miRs of miR-34a, miR-15a and let-7a into HeyA8 cells, the potency of these three miRNAs to reduce cell growth mimicked the toxicity of their 6mer seed containing siRNAs (**Figure 3B**). This suggested that a large part of their toxicity comes from the composition of the seed position 2-7. The most toxic seed in a major tumor suppressive miRNA was present in miR-34a/c, a master regulators of tumor suppression (Agostini & Knight, 2014). We directly compared the toxicity of pre-miR-34a and miR-34a^{Seed} in the same assay (**Figure 3C**). Strikingly, the toxicity evoked by these two RNA species were virtually identical. Cells showed the typical morphology we found in cells dying from toxic siRNAs (**Figure 3D**) (Hadji et al., 2014; M. Patel & Peter, 2017; Putzbach, Gao, Patel, van Dongen, et al., 2017). To determine the contribution of the 6mer seed sequence of miR-34a to its toxicity and the mode of cell death, we performed a RNA-Seq analysis on HeyA8 cells transfected with either pre-miR-34a or just its toxic seed in the neutral scaffold (miR-34a^{Seed}) (**Figure 3E**, top left) (**Table S3**). While pre-miR-34a targeted a set of genes that were not affected by miR-34a^{Seed}, the majority of genes (>72%) were targeted by both the premiR and the 6mer seed duplex (**Figure 3E**, bottom left). Both duplexes caused a similar and highly effective downregulation of the mRNAs that carry a matching 6mer seed match (**Figure 3F**) resulting in a very similar loss of survival genes when compared to set of genes not involved in cell survival (**Figure 3G**). Finally, the genes downregulated by both the premiR and the 6mer seed construct were highly enriched in genes involved in regulation of cell cycle, cell division, DNA repair, and nucleosome assembly (**Figure 3E**, right), the same GO terms that we found enriched in downregulated genes in cells dying after transfection with siRNAs containing toxic 6mer seeds (Putzbach, Gao, Patel, van Dongen, et al., 2017). These data suggest that miR-34a kills cancer cells using its toxic 6mer seed. While optimal miRNA targeting requires at least a 7mer seed interaction and also involves nucleotides at positions 13-16 of the miRNA (Grimson et al., 2007), the cell death inducing activity of this tumor suppressive miRNA may only require the 6mer seed.

Toxic 6mer Seeds Determine miRNA Strand Selection

Toxic 6mer seeds may be a driving force in miRNA evolution, whereby toxic seed sequences are either selected against - because they contribute to cell toxicity - or are preserved to operate as tumor suppressors. Based on the composition of toxic 6mer seeds and the enrichment of GC-rich motifs in survival genes, we could now ask if and when miRNAs that contain toxic G-rich sequences in positions 2-7 of their seeds evolved. When comparing all miRNAs annotated in TargetScan Human 7, we noticed that miRNAs in highly conserved miRNA seed families contained seed sequences that were much less toxic in our screen than seeds in poorly conserved miRNAs (**Figure 4A,** left panel). Consistent with our analysis on the effect of nucleotide content on seed toxicity, the 6mer seed of poorly conserved miRNA seed families had a balanced nucleotide composition in all 6 seed positions with a slight excess of G (**Figure S5**, far left). In contrast, many miRNA seed families that were highly conserved from humans to zebrafish had replaced G in the first three seed positions with the nontoxic A (**Figure S5**, far right). Interestingly, the loss of G in these positions is consistent with the asymmetry we found with G being more toxic when positioned towards the 5' end of the seed (see **Figure 2A**). In addition, highly conserved miRNA seed families somewhat avoid either G or C in position six. Weakly conserved miRNA seed families would be expected to be younger in evolutionary age than highly conserved ones. Consistent with this assumption we found that the 6mer seeds of younger miRNAs (<10 million years old) were more likely to be toxic to cells than the ones of older miRNAs (>800 million years old) (V. D. Patel & Capra,

2017) (**Figure 4A**, right panel). Most importantly, when comparing miRNAs of different ages, it became apparent that seeds of miRNAs over the last 800 million years were gradually depleted of G beginning at the 5' end and eventually also affecting positions 3-5 until the oldest ones, where G was no longer the most abundant nucleotide in any of the six positions (**Figure 4B**). It was replaced by A and U. These analyses indicated that miRNAs that are highly conserved and highly expressed in cells avoid G in potentially toxic seed positions.

miRNAs are expressed as pre-miRs and usually only one major species of mature miRNA (either the 5p or the 3p arm) is detectable in cells produced from one of the two strands of the premiR stem (Meijer, Smith, & Bushell, 2014). Consistent with the assumption that cells cannot tolerate toxic 6mer seeds, we now found that across the 780 miRNAs which have been shown to give rise to a 3p and a 5p arm, the more highly expressed arm is predicted to be significantly less toxic than the lesser expressed one (**Figure 4C**). This not strongly suggested that miRNAs with toxic seeds exist and are not readily expressed but it also allowed us to provide a new explanation for miRNA strand selection. This analysis, however, did not consider the possibility that certain miRNAs such as miR-34a may have evolved to kill cancer cells using a toxic seed. In these cases, one would actually expect the dominant arm to carry the more toxic seed. To test this assumption, we ranked all 780 miRNAs according to their ratio of the toxicity of the 6mer seed present in the predominant arm to the toxicity of the seed present in the lesser arm (**Table S4**). When we labeled the major tumor suppressive and oncogenic miRNAs, we noticed that the highly expressed arm of most of the oncogenic miRNAs contains a 6mer seed that was not toxic in our screen (**Figure 4D**, green dots). In contrast, for almost all tumor suppressive miRNAs the dominant arm contained a seed much more toxic than the lesser arm (**Figure 4D**, red dots). The overall difference in ratio between the two groups of miRNAs was highly significant. A more detailed analysis of these data revealed that the three oncogenic miRNAs with the highest ratio in toxicity between their arms, miR-363, miR-92a-2, and miR-25 were almost exclusively expressed as the nontoxic 3p form (**Figure 4E**, top). In contrast, the dominant arm of the three tumor suppressive miRNAs miR-34a, miR-34c, and miR-449b contained the most toxic seed sequence (**Figure 4E**, bottom). Interestingly, miR-449b has the same seed sequence as miR-34a and has been suggested to act as a backup miRNA for miR-34a (Concepcion, Han, et al., 2012). These data are consistent with most tumor suppressive miRNAs using the 6mer seed toxicity to kill cancer cells and suggest that this mechanism developed hundreds of millions of years ago.

DISCUSSION

The Toxic Seed Screen

We recently postulated the existence of small RNAi-active sequences that are highly toxic to cancer cells (Putzbach, Gao, Patel, Haluck-Kangas, et al., 2017). However, identifying all of them in a screen is not feasible, as a 19mer duplex siRNA would have 270 billion different sequence combinations. While miRNAs effectively target through stable 7mer seed pairing (16,348 combinations), effective targeting also involves 3' sequences in the guide strand (Grimson et al., 2007), again precluding a systematic screen for toxic sequence elements. We previously discovered that a fundamental cell type and species-independent form of seed sequence specific toxicity is determined by a 6mer seed sequence in si-/shRNAs (Putzbach, Gao, Patel, van Dongen, et al., 2017). To test the activity of any seed, we recently modified a neutral nontoxic 19mer siRNA backbone in a way that allows testing of only the guide strand by adding two 2'-O-methylation groups to positions 1 and 2 of the passenger strand (Murmann et al., 2018). This blocked loading of the passenger strand into the RISC and has now allowed us to test all 4096 possible 6mer seed sequences in a neutral siRNA duplex scaffold to determine the rules that govern 6mer seed toxicity. By ranking the 4096 6mer seeds from most toxic to least toxic, we could determine the most toxic seed composition for human and mouse cells. Our data now allow to predict the 6mer seed toxicity of any siRNA, shRNA miRNA (http://6merdb.org).

Sequence specific RNAi toxicity has been reported before (Fedorov et al., 2006; Petri & Meister, 2013). However, most of these studies involved either testing all possible si/shRNAs targeting one specific gene (Birmingham et al., 2006; Jackson et al., 2003) or using genome-wide siRNA libraries (Karlas et al., 2016; Mohr, Smith, Shamu, Neumuller, & Perrimon, 2014; Whitehurst et al., 2007). These studies were limited by the nucleotide sequences of the targeted genes or the composition of the siRNA libraries. Almost all commercial siRNA libraries are designed using specific rules that ensure efficient and specific knockdown (Bramsen et al.,

2009). They lack GC-rich sequences in their seeds as these were found to be ineffective in mediating RNAi (Gu et al., 2014; Ui-Tei et al., 2004). Our comprehensive screen now allows us to assign to all si- and shRNA a toxicity score and we provide evidence that this score can also be applied to miRNAs.

The Rules of Targeting

Our comprehensive and strand-specific analysis revealed that G-rich seeds are much more toxic than C-rich seeds, followed by U-rich seeds and no detectable toxicity in highly A-rich seeds. The preference for G richness in highly toxic 6mer seeds was confirmed with four different cell lines representing different species, cancers, and tissues of origin. Interestingly, the toxicity of a number of tumor suppressive miRNAs could also be predicted solely on the basis of their 6mer seed sequences, suggesting the rules of this toxicity are not limited to si/shRNAs. When analyzing the most toxic seeds, we found an asymmetry of high G content towards the 5' end of the seed and high C content in position 6 of the seed. The enrichment of G in the first 2-3 positions of the seed is consistent with the way Ago proteins scan mRNAs as targets. This involves mainly the first few nucleotides (positions 1-3) of the seed (Chandradoss, Schirle, Szczepaniak, MacRae, & Joo, 2015). Using a sequence matrix based on the most and least toxic 6mer seeds, we scanned 3'UTRs of survival and nonsurvival genes. We found that survival genes are enriched in CG-rich regions that are close to the 3'UTR start when compared to nonsurvival genes or when using the nontoxic AU-rich matrix.

Toxic 6mer Seeds Used by Tumor Suppressive miRNAs

When miRNAs were discovered to play a role in cancer, they were categorized as tumor-promoting and tumor-suppressing miRNAs (Esquela-Kerscher & Slack, 2006). While many miRNAs have multiple, and sometimes contradictory, activities, depending on the nature of their targets in a given cell, a number of miRNAs act predominantly oncogenic or tumor suppressive in most cancers. miR-34a is arguably the most tumor suppressive miRNA as it has been shown to be toxic to most cancer cells (Hermeking, 2010). Our strand-specific 6mer seed screen and RNA-Seq analysis revealed that the seed of miR-34a-5p accounts for most of miR-34a's toxicity to cancer cells.

Traditionally, scientists have attempted to explain tumor-promoting or tumor-suppressive activities of miRNAs by identifying targets that are either tumor-suppressive or oncogenic, respectively (Esquela-Kerscher & Slack, 2006). Examples of targets of tumor-suppressive miRNAs are the oncogenes Bcl-2 for miR-15/16 (Balatti, Pekarky, Rizzotto, & Croce, 2013) and miR-34a or c-Myc for miR-34a (Slabakova, Culig, Remsik, & Soucek, 2017). However, particularly for miR-34a, the major oncogenic targets have never been identified. Intense studies have resulted in a bewildering list of potential targets (summarized in (Slabakova et al., 2017)). Over 700 targets implicated in cancer cell proliferation, survival, and resistance to therapy have been described (Slabakova et al., 2017). Our data now suggest that certain tumor-suppressive miRNAs such as miR-34a use 6mer seed toxicity to target hundreds of housekeeping genes. While miRNAs have always been viewed as targeting entire networks of genes, they are still being studied by identifying single important targets. Our data now provide the means to rationally design new miRNA mimicking anti-cancer reagents that attack networks of survival genes.

miR-34a and Its Clinical Utility

Two important properties of miR-34a are consistent with it exerting its tumor suppressive activity through its toxic 6mer seed:

1) miR-34a is a conserved miRNA first discovered in *C. elegans* (Yang et al., 2013). In humans, miR-34a is highly expressed in many tissues (http://mirnamap.mbc.nctu.edu.tw). Consistently, miR-34a exhibits low toxicity to normal cells *in vitro* and *in vivo* (Di Martino et al., 2012). If one assumes most of its toxicity comes from its toxic 6mer seed sequence, these reports are consistent with our observation that mice treated with toxic siRNAs while slowing down tumor growth showed no signs of toxicity to normal tissues (Murmann et al., 2017). In contrast, in cancer cells miR-34a,b,c become highly expressed after genotoxic stress as they are p53-regulated (He et al., 2007).

2) miR-34a was reported to be especially active in killing CSCs (summarized in (Agostini & Knight, 2014)). This activity is consistent with our data that demonstrated that toxic 6mer seed containing siRNAs predominantly affect CSCs (Ceppi et al., 2014).

In 2013, miR-34a (MRX34) became the first miRNA to be tested in a phase I clinical trial of unresectable primary liver cancer (Beg et al., 2017). The study was recently terminated and reported immune-related adverse effects in several individuals. It was suggested that these adverse effects may have been caused by either a reaction to the liposome-based carrier or the use of double-stranded RNA (Slabakova et al., 2017). In addition, they may be due to an undesired gene modulation by miR-34a itself. Alternative forms of delivery were suggested to possibly improve this negative outcome (Slabakova et al., 2017).

We now demonstrate that miR-34a miRNA exerting toxicity mostly through its 6mer seed. This suggests that its 700 known targets may be part of the network of survival genes that are targeted. Our data would suggest to analyze the treated patients to determine whether the toxicity seen was due to targets specific to pre-miR-34a in noncancer cells rather than the toxic miR-34a 6mer seed. The comparison of the RNA-Seq data of cells treated with either pre-miR-34a or miR-34^{Seed} now allows to determine whether these two activities can be separated.

The Evolution of miRNAs

It was shown before that miRNAs overall avoid seed sequences that target the 3‘UTR of survival/housekeeping genes (Stark, Brennecke, Bushati, Russell, & Cohen, 2005; Zare, Khodursky, & Sartorelli, 2014). Survival genes therefore are depleted in seed matches for the most abundant miRNAs in a cell. That also means 3‘UTRs of survival genes must be enriched in sequences not targeted by the seeds present in most miRNAs. Our combined data now suggest it is these sequences that toxic siRNAs and tumor suppressive miRNAs with toxic 6mer seeds are targeting. Our analyses also suggest that most miRNAs have evolved over the last 800 million years by gradually depleting G in their seeds beginning with the 5' end (see Figure 4B). While evolution has selected to reduce the number of seed matches in the 3‘UTR of survival genes for abundant miRNAs, we realized that every cell has a powerful suicide mechanism built in: the production of small RNAi-active sequences that target the sites that are avoided by most miRNAs. Expressing miRNAs with such toxic seeds could be an effective way to eliminate unwanted cells. After characterizing the most toxic 6mer seeds, we identified miRNAs that mediate toxicity through inducing this death mechanism. In addition, we discovered a rationale for the strand selectivity of many miRNAs. All miRNAs are processed from the pre-miR stem loop structure, and the mature miRNA, which acts as guide strand once loaded into the RISC, can be derived from either the 5' (5p) and in the 3' arm (3p), while the other arm is degraded. Our data now suggest that during evolution, the most abundant miRNAs have evolved to use the arm with the lower 6mer seed toxicity as the active guide strand, presumably to avoid killing cells. In most but not all cases this is the 3p arm (data not shown). The predominant strand contains the toxic seed in only a minority of tumor-suppressive miRNAs. In fact, just by ranking miRNAs according to whether they express the arm with the seed of higher toxicity, we could separate established oncogenic from tumor suppressive miRNAs (see Figure 4D). Using this method, it is now possible to identify novel tumor suppressive miRNAs (see 6merdb.org).

In summary, we have determined the rules of RNAi targeting by toxic 6mer seeds. These rules allowed us to predict with some certainty which si/shRNA sequence and which miRNA has the potential to kill cells through their toxic seed. Interestingly, toxic 6mer seeds are present in a number of tumor-suppressive miRNAs that can kill cancer cells. Our data also provide new insights into the evolution of miRNAs and provide a new rationale to explain miRNA strand selection.

METHODS

Reagents and Antibodies

HeyA8 (RRID:CVCL_8878) and H460 (ATCC HTB-177) cells were cultured in RPMI1640 medium (Cellgro Cat#10-040) supplemented with 10% FBS (Sigma Cat#14009C) and 1% L-Glutamine (Corning Cat#25-005). 3LL cells (ATCC CRL-1642) were cultured in DMEM medium (Gibco Cat#12430054) supplemented with 10% FBS and 1% L-Glutamine. Mouse hepatocellular carcinoma cells M565 cells were described previously (23)

and cultured in DMEM/F12 (Gibco Cat#11330) supplemented with 10% FBS, 1% L-Glutamine and ITS (Corning #25-800-CR). Anti-Argonaute-2 antibody (cat#ab186733) was purchased from Abcam, the anti-β-actin antibody from Santa Cruz (#sc-47778).

siRNA Screens and Cell Viability Assay

The nontoxic siRNA backbone used in the 4096 screen was designed as previously described (26). Briefly, the siNT2 sequence was used as a starting point and four positions in the center of siNT2 were replaced with the complementary nucleotides in order to remove any identity between the backbone siRNA and the toxic siL3 while retaining the same GC content. Two 2'-O-methylation groups were added to positions 1 and 2 of the passenger strand. The 6mer seed region (position 2-7 on the guide strand) were then replaced with one of the 4096 possible seeds. Transfection efficiency was optimized for each of the four cell line individually. For HeyA8 and M565 cells, the reverse transfection and the quantification of cellular ATP content in the 4096 seed duplexes screen was done as previously described (26). For 3LL and H460 cells, the repeat screen with the 76 duplexes were done in a very similar manner except that the amount of RNAiMAX and the number of cells plated. For 3LL cells, 150 cells per well in 384 well plate was added and 9.3 μl of Lipofectamine RNAiMax was mixed with 990.7 μl of Opti-MEM. For H460 cells, 420 cells per well in 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 an IncuCyte experiment, 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 AGO2 knockdown experiment, 100,000 cells/well HeyA8 cells were reverse-transfected in six-well plate 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 Table S1) at 10 nM and monitored in the IncuCyte Zoom. 0.1 μl/well RNAiMAX was used. 48 hours after transfection, cells were lysed in RIPA buffer for western blot analysis as described previously in (22).

All custom siRNA oligonucleotides were ordered from integrated DNA technologies (IDT) and annealed according to the manufacturer's instructions. The following siRNA sequences were used:
siNT1 sense: rUrGrGrUrUrUrArCrArUrGrUrCrGrArCrUrArATT;
siNT1 antisense: rUrUrArGrUrCrGrArCrArUrGrUrArArArCrCrAAA;
siNT2 sense: rUrGrGrUrUrUrArCrArUrGrUrUrGrUrGrUrGrATT;
siNT2 antisense: rUrCrArCrArCrArArCrArUrGrUrArArArCrCrAAA;
si-miR-34a^{Seed} sense: mUmGrGrUrUrUrArCrArUrGrUrArCrUrGrCrCrATT;
si-miR-34a^{Seed} antisense: rUrGrGrCrArGrUrArCrArUrGrUrArArArCrCrAAA;
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-15a-5p (Ambion, Cat. No# PM10235), and miRNA precursor negative control #1 (Ambion, Cat. No# AM17110).

Monitoring Growth over Time by IncuCyte

To monitor cell growth over time, 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).

RNA-Seq Analysis

For RNA-Seq data in Figure 3E 50,000 cells/well HeyA8 cells were reversed transfected in 6-well plates with 10 nM of either pre-miR-34a or si-miR-34a^{Seed} with their respective controls. The transfection mix was replaced 24 hours after transfection. The cells were lysed in QIAzol Lysis reagent (Qiagen, Cat. No# 79306) 48 hours

after transfection. 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 in (22) at the Genomics Core facility at the University of Chicago. Reads were aligned to the hg38 version of the human genome, using STAR v2.5.2. 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.

Data Analyses

GSEA was performed using the GSEA software version 3.0 from the Broad Institute downloaded from https://software.broadinstitute.org/gsea/. A ranked list was generated by sorting genes according the Log_{10}(Fold Downregulation). The Pre-ranked function was used to perform GSEA using the ranked list. 1000 permutations were used. Default settings were used. The ~1800 survival genes and ~420 nonsurvival genes defined previously (Putzbach, Gao, Patel, van Dongen, et al., 2017) were used as custom gene sets. Default settings were used.

Sylamer analysis (van Dongen, Abreu-Goodger, & Enright, 2008) was used to find enrichment of small word motifs in the 3'UTRs of genes enriched in those that are most downregulated. A list of 3'UTRs was generated by eliminating repetitive regions as described previously (Putzbach, Gao, Patel, van Dongen, et al., 2017) to use for the Sylamer analysis. Default settings were used. Bonferroni adjusted p-values were calculated by multiplying the unadjusted p-values by the number of permutations for each length of word searched for.

The GO enrichment analyses shown in Figures 2B and Figure S3A were performed using the GOrilla GO analysis tool at http://cbl-gorilla.cs.technion.ac.il using default setting and a p-value cut-off of 10^{-11}. The GO enrichment analyses shown in Figure 3E was done using the DAVID GO tool (v. 6.8) at https://david.ncifcrf.gov/home.jsp.

Plots showing the contribution of the four nucleotides G, C, A and U at each of the 6mer seed positions were generated using the Weblogo tool at http://weblogo.berkeley.edu/logo.cgi.

The Ingenuity Pathway Analysis (IPA) software from QIAGEN was used to identify the pathways enriched in the top 4288 genes with the highest PWM scores (>400). A list of the 4288 genes was used as the input for the core analysis provided by the IPA. The two miRNA networks shown in Figure S4 were identified as the top two Top Regulator Effect Networks by the core analysis.

All probability density plots comparing the distribution of either the nucleotide content or position of the first best match in the 3'UTR were generated using the R package ggplot2 (Figure 2C, Figure 2D, Figure S3B, and Figure S3C). The peaks for each probability distribution in the density plots were also calculated in R.

The PWMScore tool scores nucleotide sequences based on matches to a sequence motif represented by a position weight matrix (PWM) or a base probability matrix. Two scoring methods can be applied: i) the 'sum occupancy score', which sums over all motif matches weighted by their respective strengths or ii) the 'best single match scoring' mode, which reports the score of the best single match within the sequence, together with the location of all best hits. More details about the arithmetic of score computation can be found on the Web server (http://ccg.vital-it.ch/pwmtools/pwmscore.php). Here, all human 3'UTRs were scored for the best single match in the forward direction, using a PWM that represents the nucleotide composition of the 20 most toxic seeds.

To compare the nucleotide content of the 3'UTR of genes with the highest versus the lowest PWM scores, the 3'UTR of all human genes were extracted as previously described (Putzbach, Gao, Patel, van Dongen, et al., 2017) and nucleotide content was calculated using an Excel script. The 4288 genes with the highest PWM scores (>400) and the 4834 genes with the lowest scores (<90) were then extracted. The whole set of genes used in the PWM analysis (18,795 genes total) was used as the background control (Figure S3B). Similarly, to compare the nucleotide content of the 3'UTR of survival genes versus nonsurvival genes, 1840 survival genes (out of 1882) and 411 nonsurvival genes (out of 423) that were included in the PWM analysis (Figure 2C).

To compare the position of the first best match for the toxic versus non toxic matrix in the 3'UTRs, only genes with the highest PWM scores for either the toxic matrix (4288 genes) or the nontoxic matrix (8162 genes)

were considered. To examine the distribution of the first best match position within the first 1000 bp after the start of 3'UTRs, 2623 genes in the toxic matrix (out of 4288) and 4902 genes (out of 8162) that had their first best match position within the first 1000 bp were included in the density plot in Figure 2D. To exclude a confounding effects of the 3'UTR length, genes with 3'UTR length less than 1000 bp were eliminated in the next analysis which included 1803 genes in the toxic matrix and 3627 genes in the non toxic matrix (Figure 2D, right panel). The distribution of the positions of the best first match in either the toxic matrix or the nontoxic 6mer seed matrix in the 3'UTRs of survival genes or nonsurvival genes were done in a similar fashion. For the toxic matrix, 574 survival genes and 131 nonsurvival genes with the first best match position within the first 1000 bp and with 3'UTR length longer than 1000 bp were plotted (Figure S3C, left). For the nontoxic matrix, 618 survival genes and 132 nonsurvival genes with the first best match position within the first 1000 bp and with 3'UTR length longer than 1000 bp were plotted (Figure S3C, right).

Relation between miRNA Seed Conservation and Age and Toxicity

Information on miRNA seed family conservation and seed sequence were downloaded at http://www.targetscan.org/vert_71/ from TargetScan Human 7.1. The toxicity of each mature human miRNA arm sequence in the TargetScan dataset was assigned according to the toxicity induced by the siRNA in HeyA8 siRNA screen harboring the identical 6mer seed sequence. A list of miRNA ages corresponding to ~2500 microRNA loci was acquired from Patel and Capra and was calculated using a modified version of ProteinHistorian (V. D. Patel & Capra, 2017). This list was used to assign ages to roughly 1400 mature miRNA arms found in the TargetScan dataset.

TargetScan 7.1 partitions the seed family conservation into four groups: highly conserved (group #2), conserved (group #1), low conservation but still annotated as a miRNA (group #0), and low conservation with the possibility of misannotation (group #-1). Probability density plots for the assigned seed-dependent toxicity were generated for each seed family conservation group using ggplot2 in Rstudio. Probability density plots were also generated to show how young (<10 million years) and old (>800 million years) miRNAs compare in terms of the seed-dependent toxicity. Further partitioning of the miRNAs according to their predicted age was done so we could analyze how age affects the nucleotide composition of the seed. We also analyzed how seed compositions differed between different seed family conservation groups. Nucleotide composition was visualized using probability density plots. Nucleotide composition at each individual position in the seed was also assessed. All differences between groups in terms of seed-dependent toxicity and nucleotide composition in the seed were analyzed using a two-sample two-tailed Kolmogorov-Smirnov test. Comparison of seed-dependent toxicity was also made between the predominantly- and lesser-expressed arms by assigning toxicity, according to the HeyA8 cell screen, to human miRNA-1 to 500, plus let-7. The miRNA genes from this list were then assigned as either tumor suppressive or oncogenic based on the literature and then ranked from most to least toxic seed sequence (see Figure 4D).

Assessment of Dominant Arm Seed Toxicity

The expression of miRNA 5p and 3p arms across 135 tissue samples were collected from MIRMINE (Panwar, Omenn, & Guan, 2017). A miRNA was considered expressed if the sum of the normalized reads for both arms was above 5 for each sample. A miRNA arm was considered the dominant species if its expression was at least 25% greater than the other arm per sample. The dominant arm for each miRNA across all samples was calculated by determining which arm was dominantly expressed in more samples. The miRNAs that have only one annotated arm in miRBase were considered to have only one dominant arm. The seed toxicity values for each arm were extracted from the 4096 siRNA screen data (% viability) for the human HeyA8 cells and used to calculate the ratio between the dominant arm's toxicity and the lesser arm's toxicity; miRNAs that only had one expressed arm were not considered in this analysis but are all included in the website: 6merdb.org.

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. 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 (Figures S1B, S1C, S2A) 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 in Figure 4D.

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). All 6mer seed toxicity data of the 4096 siRNA screen in HeyA8 and M565 cells are available in searchable form at http://6merdb.org.

ACKNOWLEDGEMENTS

We are indebted to Dr. Leon Platanias for his generous support of the seed screen and to Denise Scholtens for help with biostatistics. This work was funded by training grant T32CA009560 (to W.P.) and R35CA197450 (to M.E.P.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.E.P.; Investigation, Q.Q.G., W.E.P., A.E.M., S.C., J.M.P., G.A., and E.T.B.; Writing-Original draft, M.E.P.; Writing-Review & Editing, Q.Q.G, W.E.P, and M.E.P.; Supervision, M.E.P.

DECLARATION OF INTEREST

The authors declare no competing interests.

Figure legends:

Figure 1: A comprehensive screen to identify the most toxic 6mer seeds.

(**A**) Schematic of the siRNA backbone used in the 4096 seed duplexes toxicity screen. 2'-O-methylation modifications at position 1 and 2 of the passenger strand were added to prevent passenger strand loading (marked by red Xs) (Murmann et al., 2018). The nucleotides that remained constant in all the duplexes are in blue. The variable 6mer seed sequence is in red and boxed in blue.

(**B**) Results of the 4096 6mer seed duplex screen in HeyA8 cells. Cells were reverse transfected in triplicates in 384 well plates with 10 nM of individual siRNA. The cell viability of each 6mer seed duplex was determined by quantifying cellular ATP content 96 hours after transfection. All 4096 6mer seeds are ranked by their effects on cell viability from the most toxic (left) to the least toxic (right). Rankings of the 6mer seeds of four previously characterized CD95L derived siRNAs (siL1, siL2, siL3, and siL4) are highlighted in red. The top 200 toxic seeds (cell viability <10%) are highlighted in pink and the bottom 200 least toxic seeds (cell viability >100%) are highlighted in green.

(**C**) Cell viability of the 19 seed duplexes with the highest content (>80%) in the 6mer seed region for each nucleotide group in two human and two mouse cell lines: HeyA8 (first row), H460 (second row), 3LL (third row), and M565 (last row). The G-rich seeds are in red, the C-rich seeds are in orange, the A-rich seeds are in blue, and the U-rich seeds are in green. The purple box highlights an example of reduced toxicity in all four cell lines (grey stippled line) caused by moving an A from the 5' to the 3' end of the 6mer seed. p-values between groups of duplexes were calculated using Student's ttest.

(**D**) Nucleotide composition at each of the 6 seed positions in the top 200 most toxic (left) or the top 200 least toxic (right) seed duplexes in either HeyA8 cells (left panels) or M565 cells (right panels).

Figure 2: 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 the top 20 most toxic seeds. The reverse complement of the toxic matrix was used as the input for the PWMScore analysis to search for potential targets of the toxic seeds.

(**B**) Results of a GOrilla gene ontology analysis using a gene list ranked by the PWM matrix match score with the toxic matrix in their 3’UTRs ranked from the highest to the lowest score. Only GO clusters that had a significance of enrichment of at least 10^{-11} are shown.

(**C**) Comparison of nucleotide content between a group of ~1800 genes critical for survival (SGs) and a control set of ~400 genes that are not required for survival (nonSGs) (Putzbach, Gao, Patel, van Dongen, et al., 2017) (right panel). p-values were calculated using a two-sample two-sided K-S test comparing the density distribution of SGs and nonSGs. Peak maxima are given.

(**D**) Distribution of the location of the best first match in target sites of either the toxic or the nontoxic 6mer seed matrix in the 3’UTRs. Left: the best first match to the matrix is shown for all 3‘UTRs of genes with the highest PMW Score (2623 genes with the toxic matrix and 4902 genes with the nontoxic matrix) regardless of length. Only the best first match positions that are within first 1000 bases were analyzed. Right: The same analysis on the left but only genes with 3'UTRs 1 kb or larger were analyzed (1803 genes with the toxic matrix and 3627 genes with the nontoxic matrix). Peak maxima are given. p-values were determined using the K-S test.

(**E**) Contribution of synonymous single nucleotide mutations that result in a loss of Cs (blue column) to all documented point mutations (white column) across 27 human cancers ranked according to their mutational load.

Figure 3: Tumor suppressive miRNAs inhibit cancer cell growth via toxic 6mer seeds.

(**A**) All 4096 6mer seeds ranked from the lowest viability (highest toxicity) to the highest viability (lowest toxicity) in HeyA8 cells. Locations of 6mer seeds present in major tumor suppressive (red) or tumor promoting (green) miRNAs are highlighted as individual bars.

(**B**) Percent cell confluence over time of HeyA8 cells transfected with 5 nM of either tumor suppressive miRNA precursors including pre-miR-let-7a, pre-miR-15a, and pre-miR-34a or a miRNA precursor nontargeting control. *Two-way ANOVA p-value between cells treated with pre-miR-(NC) and pre-let-7a is 0.0.

(**C**) Percent cell confluence over time of HeyA8 parental cells transfected with either pre-miR-34a or si-miR-34a^{Seed} and compared to their respective controls (pre-miR (NC) for pre-miR-34a and siNT2 for si-miR-34a^{Seed}) at 10 nM.

(**D**) Morphology observed in HeyA8 cells transfected with either pre-miR-34a or si-miR-34a^{Seed} compared to their respective controls at 10 nM three days after transfection.

(**E**) *Top left*: Alignment of the sequences of miR-34a-5p and miR-34a^{Seed} with the 6mer highlighted. *Bottom left:* Overlap of RNAs detected by RNA-Seq downregulated in HeyA8 cells 48 hrs after transfection with either miR-34^{Seed} or pre-miR-34a when compared to either siNT1 or a nontargeting pre-miR, respectively. Right: DAVID GO analysis of the two indicated groups. The top 10 most enriched GO terms that reach statistical significance are shown.

(**F**) Sylamer plots for the list of 3‘UTRs of mRNAs in cells treated with either pre-miR-34a (top) or miR-34^{Seed} (bottom) ordered from down-regulated to up-regulated. The most highly enriched sequence is shown which in each case is the 6mer seed match of the introduce 6mer seed. Bonferroni-adjusted p-values are shown.

(**G**) Gene set enrichment analysis for a group of 1846 survival genes (top 4 panels) and 416 nonsurvival genes (bottom 2 panels) identified in a genome-wide CRISPR lethality screen (T. Wang et al., 2015) after transfecting HeyA8 cells with either pre-miR-34a or mir-34a^{Seed}. siNT1 and a nontargeting premiR served as controls, respectively. p-values indicate the significance of enrichment.

Figure 4: 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 miRNAs = 2588) or of very old (>800 Million years) miRNAs or very young (<10 Million years) miRNAs (right panel, total number of miRNAs = 1025). Kolmogorov–Smirnov two-sample two-sided 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 either highly or lesser expressed arms of 780 miRNAs.
(**D**) 780 miRNAs ranked according to the ratio of viability of the seed (as determined in the seed screen) of the predominantly expressed and the lesser-expressed arm. Established oncogenic miRNAs are shown in green, tumor-suppressive miRNAs are shown in red. The predominant arm is given for each miRNA (in parenthesis). p-value of the distribution of oncogenic versus tumor suppressive miRNAs was calculated using KS 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 predominant 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, nontoxic ones in green.

REFERENCES

Agostini, M., & Knight, R. A. (2014). miR-34: from bench to bedside. *Oncotarget, 5*(4), 872-881. doi: 10.18632/oncotarget.1825

Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature, 455*(7209), 64-71.

Balatti, V., Pekarky, Y., Rizzotto, L., & Croce, C. M. (2013). miR deregulation in CLL. *Adv Exp Med Biol, 792*, 309-325. doi: 10.1007/978-1-4614-8051-8_14

Beg, M. S., Brenner, A. J., Sachdev, J., Borad, M., Kang, Y. K., Stoudemire, J., . . . Hong, D. S. (2017). Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Invest New Drugs, 35*(2), 180-188. doi: 10.1007/s10637-016-0407-y

Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature, 409*(6818), 363-366. doi: 10.1038/35053110

Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., . . . Khvorova, A. (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods, 3*(3), 199-204. doi: 10.1038/nmeth854

Bramsen, J. B., Laursen, M. B., Nielsen, A. F., Hansen, T. B., Bus, C., Langkjaer, N., . . . Kjems, J. (2009). A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res, 37*(9), 2867-2881. doi: 10.1093/nar/gkp106

Ceppi, P., Hadji, A., Kohlhapp, FJ., Pattanayak, A., Hau, A., Xia, L., . . . Peter, M. E. (2014). CD95 and CD95L promote and protect cancer stem cells. *Nature Commun, 5*, 5238.

Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J., & Joo, C. (2015). A Dynamic Search Process Underlies MicroRNA Targeting. *Cell, 162*(1), 96-107. doi: 10.1016/j.cell.2015.06.032

Concepcion, C. P., Bonetti, C., & Ventura, A. (2012). The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer J, 18*(3), 262-267. doi: 10.1097/PPO.0b013e318258b60a

Concepcion, C. P., Han, Y. C., Mu, P., Bonetti, C., Yao, E., D'Andrea, A., . . . Ventura, A. (2012). Intact p53-dependent responses in miR-34-deficient mice. *PLoS Genet, 8*(7), e1002797. doi: 10.1371/journal.pgen.1002797

Di Martino, M. T., Leone, E., Amodio, N., Foresta, U., Lionetti, M., Pitari, M. R., . . . Tassone, P. (2012). Synthetic miR-34a mimics as a novel therapeutic agent for multiple myeloma: in vitro and in vivo evidence. *Clin Cancer Res, 18*(22), 6260-6270. doi: 10.1158/1078-0432.CCR-12-1708

Esquela-Kerscher, A., & Slack, F. J. (2006). Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer, 6*(4), 259-269.

Eulalio, A., Huntzinger, E., & Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol, 15*(4), 346-353. doi: 10.1038/nsmb.1405

Fedorov, Y., Anderson, E. M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., . . . Khvorova, A. (2006). Off-target effects by siRNA can induce toxic phenotype. *RNA, 12*(7), 1188-1196. doi: 10.1261/rna.28106

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell, 27*(1), 91-105.

Gu, S., Zhang, Y., Jin, L., Huang, Y., Zhang, F., Bassik, M. C., . . . Kay, M. A. (2014). Weak base pairing in both seed and 3' regions reduces RNAi off-targets and enhances si/shRNA designs. *Nucleic Acids Res, 42*(19), 12169-12176. doi: 10.1093/nar/gku854

Hadji, A., Ceppi, P., Murmann, A. E., Brockway, S., Pattanayak, A., Bhinder, B., . . . Peter, M.E. (2014). Death induced by CD95 or CD95 ligand elimination. *Cell Reports, 10*, 208-222.

Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., & Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev, 18*(24), 3016-3027.

He, X., He, L., & Hannon, G. J. (2007). The guardian's little helper: microRNAs in the p53 tumor suppressor network. *Cancer Res, 67*(23), 11099-11101. doi: 10.1158/0008-5472.CAN-07-2672

Hermeking, H. (2010). The miR-34 family in cancer and apoptosis. *Cell Death Differ, 17*(2), 193-199. doi: 10.1038/cdd.2009.56

Hua, Y.J., Larsen, N., Kalyana-Sundaram, S., Kjems, J., Chinnaiyan, A. M., & Peter, M. E. (2013). miRConnect 2.0: Identification of antagonistic, oncogenic miRNA families in three human cancers. *BMC Genomics, 14*, 179.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., & Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science, 293*(5531), 834-838. doi: 10.1126/science.1062961

Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., . . . Linsley, P. S. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol, 21*(6), 635-637. doi: 10.1038/nbt831

Karlas, A., Berre, S., Couderc, T., Varjak, M., Braun, P., Meyer, M., . . . Lecuit, M. (2016). A human genome-wide loss-of-function screen identifies effective chikungunya antiviral drugs. *Nature communications, 7*, 11320. doi: 10.1038/ncomms11320

Lai, E. C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet, 30*(4), 363-364.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., . . . Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature, 499*(7457), 214-218. doi: 10.1038/nature12213

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J, 23*(20), 4051-4060.

Leuschner, P. J., Ameres, S. L., Kueng, S., & Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep, 7*(3), 314-320. doi: 10.1038/sj.embor.7400637

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell, 115*(7), 787-798.

Louie, E., Ott, J., & Majewski, J. (2003). Nucleotide frequency variation across human genes. *Genome Res, 13*(12), 2594-2601. doi: 10.1101/gr.1317703

Meijer, H. A., Smith, E. M., & Bushell, M. (2014). Regulation of miRNA strand selection: follow the leader? *Biochem Soc Trans, 42*(4), 1135-1140. doi: 10.1042/BST20140142

Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol, 3*(3), REVIEWS0004.

Mohr, S. E., Smith, J. A., Shamu, C. E., Neumuller, R. A., & Perrimon, N. (2014). RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol, 15*(9), 591-600. doi: 10.1038/nrm3860

Murmann, A. E., Gao, Q. Q., Putzbach, W. T., Patel, M., Bartom, E. T., Law, C. Y., . . . Peter, M. E. (2018). Small interfering RNAs based on huntingtin trinucleotide repeats are highly toxic to cancer cells. *EMBO Rep, 19*, e45336.

Murmann, A. E., McMahon, K. M., Halluck-Kangas, A., Ravindran, N., Patel, M., Law, C., . . . Peter, M. E. (2017). Induction of DISE in ovarian cancer cells in vivo. *Oncotarget, 8*, 84643-84658.

Panwar, B., Omenn, G. S., & Guan, Y. (2017). miRmine: a database of human miRNA expression profiles. *Bioinformatics, 33*(10), 1554-1560. doi: 10.1093/bioinformatics/btx019

Patel, M., & Peter, M. E. (2017). Identification of DISE-inducing shRNAs by monitoring cellular responses. *Cell Cycle*, doi: 10.1080/15384101.15382017.11383576. doi: 10.1080/15384101.2017.1383576

Patel, V. D., & Capra, J. A. (2017). Ancient human miRNAs are more likely to have broad functions and disease associations than young miRNAs. *BMC Genomics, 18*(1), 672. doi: 10.1186/s12864-017-4073-z

Peter, M. E. (2009). Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle, 8*(6), 843-852.

Petri, S., & Meister, G. (2013). siRNA design principles and off-target effects. *Methods Mol Biol, 986*, 59-71. doi: 10.1007/978-1-62703-311-4_4

Putzbach, W., Gao, Q. Q., Patel, M., Haluck-Kangas, A., Murmann, A. E., & Peter, M. E. (2017). DISE - A Seed Dependent RNAi Off-Target Effect that Kills Cancer Cells. *Trends in Cancer*, In press.

Putzbach, W., Gao, Q. Q., Patel, M., van Dongen, S., Haluck-Kangas, A., Sarshad, A. A., . . . Peter, M. E. (2017). Many si/shRNAs can kill cancer cells by targeting multiple survival genes through an off-target mechanism. *Elife, 6*, e29702.

Schirle, N. T., & MacRae, I. J. (2012). The crystal structure of human Argonaute2. *Science, 336*(6084), 1037-1040. doi: 10.1126/science.1221551

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature, 455*(7209), 58-63.

Slabakova, E., Culig, Z., Remsik, J., & Soucek, K. (2017). Alternative mechanisms of miR-34a regulation in cancer. *Cell death & disease, 8*(10), e3100. doi: 10.1038/cddis.2017.495

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., & Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell, 123*(6), 1133-1146. doi: 10.1016/j.cell.2005.11.023

Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., . . . Saigo, K. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res, 32*(3), 936-948. doi: 10.1093/nar/gkh247

van Dongen, S., Abreu-Goodger, C., & Enright, A. J. (2008). Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods, 5*(12), 1023-1025. doi: 10.1038/nmeth.1267

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., . . . Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science, 350*(6264), 1096-1101. doi: 10.1126/science.aac7041

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., & Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature, 456*(7219), 209-213. doi: 10.1038/nature07315

Whitehurst, A. W., Bodemann, B. O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., . . . White, M. A. (2007). Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature, 446*(7137), 815-819. doi: 10.1038/nature05697

Yang, J., Chen, D., He, Y., Melendez, A., Feng, Z., Hong, Q., . . . Chen, X. (2013). MiR-34 modulates Caenorhabditis elegans lifespan via repressing the autophagy gene atg9. *Age (Dordr), 35*(1), 11-22. doi: 10.1007/s11357-011-9324-3

Yi, R., Qin, Y., Macara, I. G., & Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev, 17*(24), 3011-3016.

Zare, H., Khodursky, A., & Sartorelli, V. (2014). An evolutionarily biased distribution of miRNA sites toward regulatory genes with high promoter-driven intrinsic transcriptional noise. *BMC Evol Biol, 14*, 74. doi: 10.1186/1471-2148-14-74

Figure 1

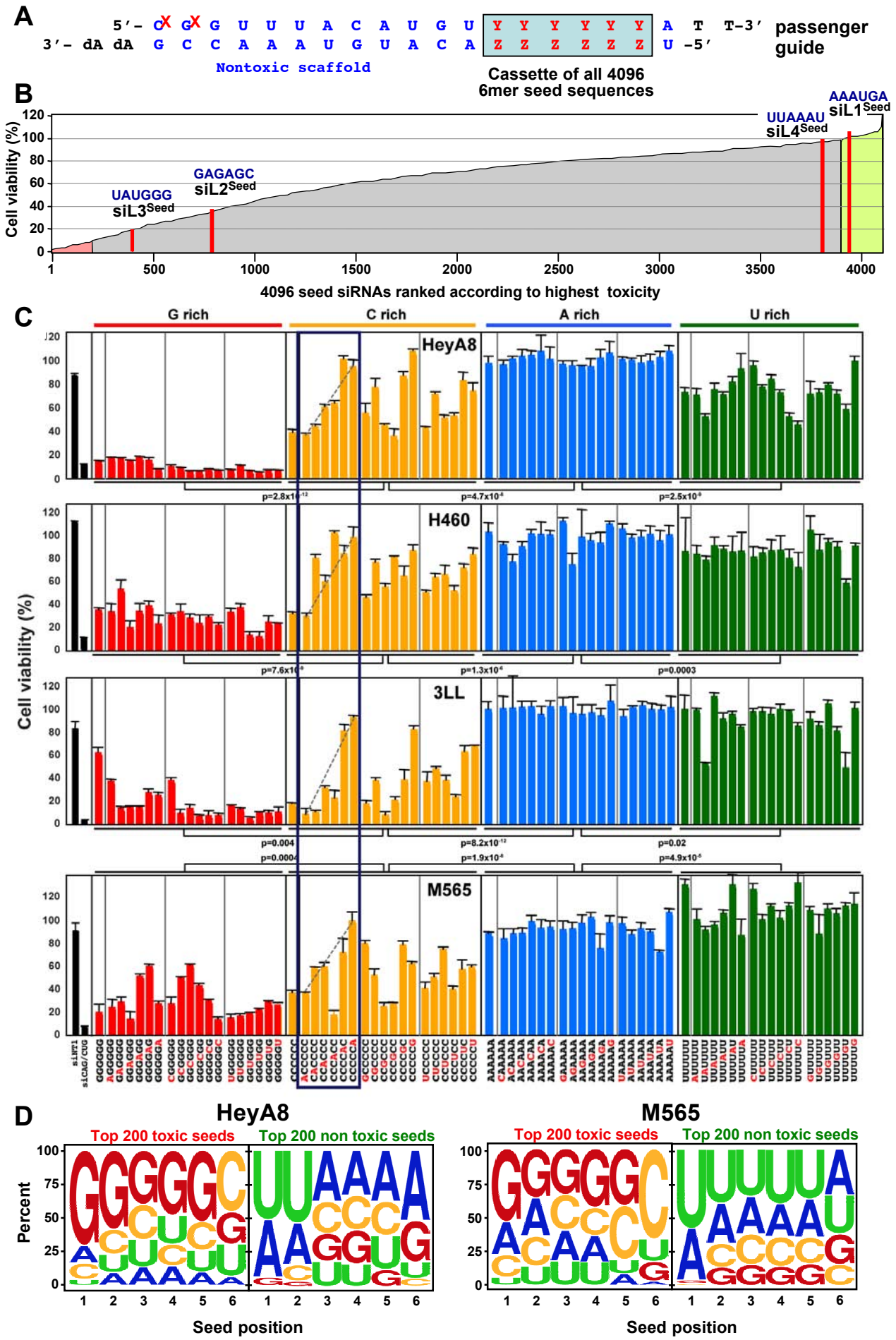
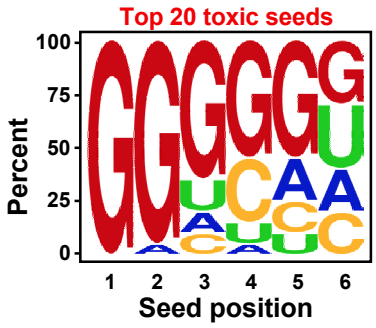

A
5′- C X G X G U U U A C A U G U Y Y Y Y Y Y A T T-3′ passenger
3′- dA dA G C C A A A U G U A C A Z Z Z Z Z Z U -5′ guide
Nontoxic scaffold
Cassette of all 4096 6mer seed sequences
B
Cell viability (%)
UAUGGG siL3Seed
GAGAGC siL2Seed
UUAAAU siL4Seed
AAAUGA siL1Seed
4096 seed siRNAs ranked according to highest toxicity
C
G rich
C rich
A rich
U rich
HeyA8
H460
3LL
M565
Cell viability (%)
D
HeyA8
M565
Top 200 toxic seeds
Top 200 non toxic seeds
Percent
Seed position

Figure 2

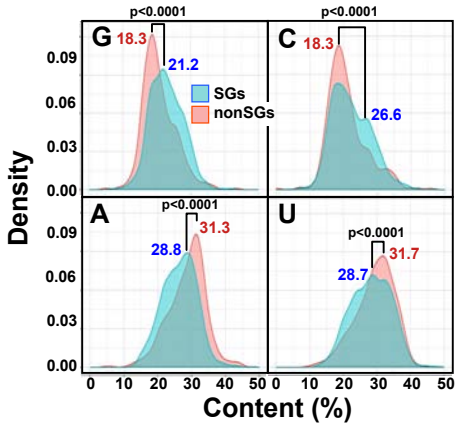
A



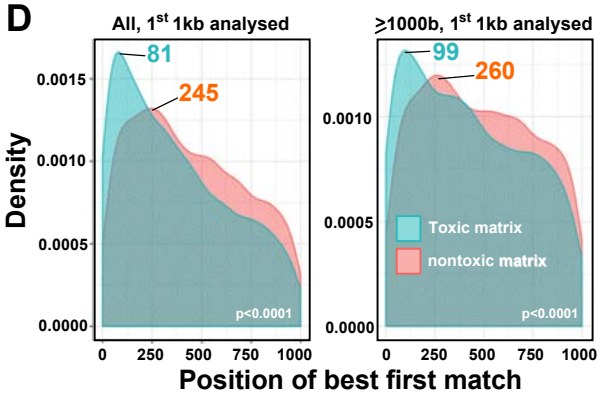
B

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0032502	developmental process	1.66E-14	2.52E-10	1.40
GO:0006357	regulation of transcription by RNA polymerase II	2.86E-14	2.17E-10	1.57
GO:0080090	regulation of primary metabolic process	3.80E-13	1.92E-09	1.25
GO:0031325	positive regulation of cellular metabolic process	6.39E-13	2.42E-09	1.39
GO:0048856	anatomical structure development	6.87E-13	2.08E-09	1.50
GO:0048522	positive regulation of cellular process	7.06E-13	1.78E-09	1.31
GO:0050794	regulation of cellular process	8.92E-13	1.93E-09	1.16
GO:0051173	positive regulation of nitrogen compound metabolic process	9.24E-13	1.75E-09	1.40
GO:0065007	biological regulation	9.67E-13	1.63E-09	1.13
GO:0010604	positive regulation of macromolecule metabolic process	1.72E-12	2.60E-09	1.39
GO:0031323	regulation of cellular metabolic process	2.21E-12	3.05E-09	1.24
GO:0050793	regulation of developmental process	3.16E-12	3.99E-09	1.44
GO:0048523	negative regulation of cellular process	3.63E-12	4.23E-09	1.30
GO:0051171	regulation of nitrogen compound metabolic process	6.14E-12	6.64E-09	1.24
GO:0009893	positive regulation of metabolic process	6.18E-12	6.24E-09	1.36
GO:0050789	regulation of biological process	6.23E-12	5.90E-09	1.14
GO:0048518	positive regulation of biological process	7.15E-12	6.37E-09	1.25
GO:0051254	positive regulation of RNA metabolic process	9.66E-12	8.13E-09	1.57

C



D



E

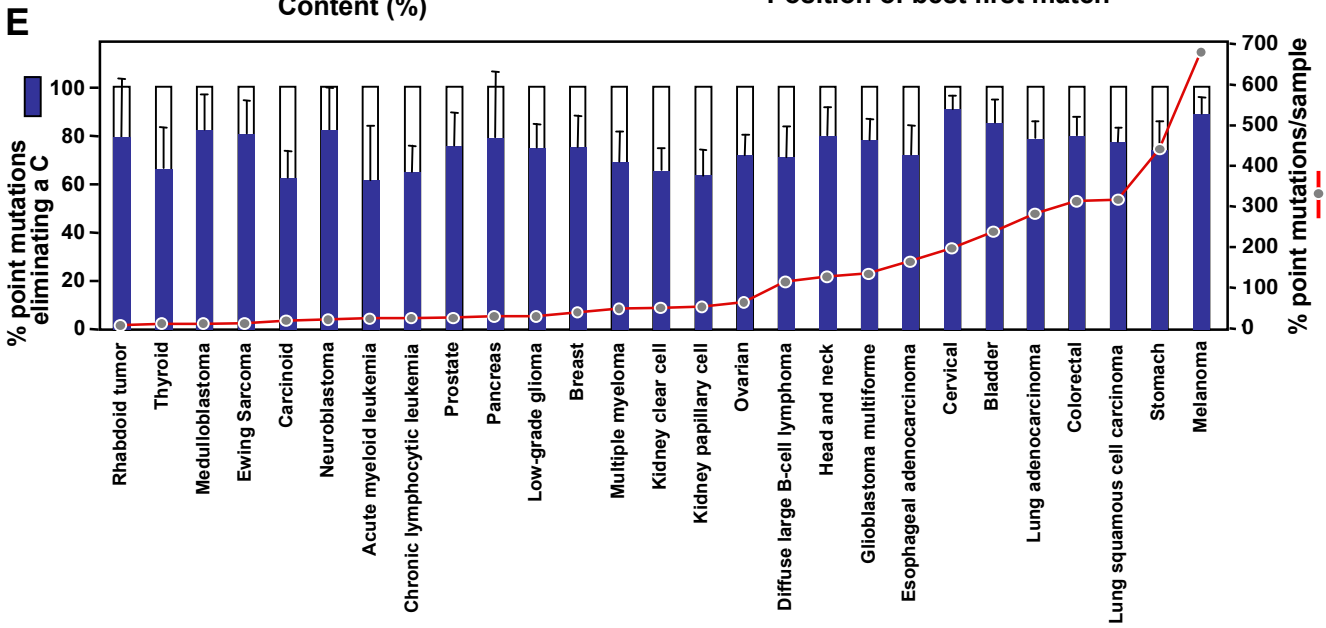
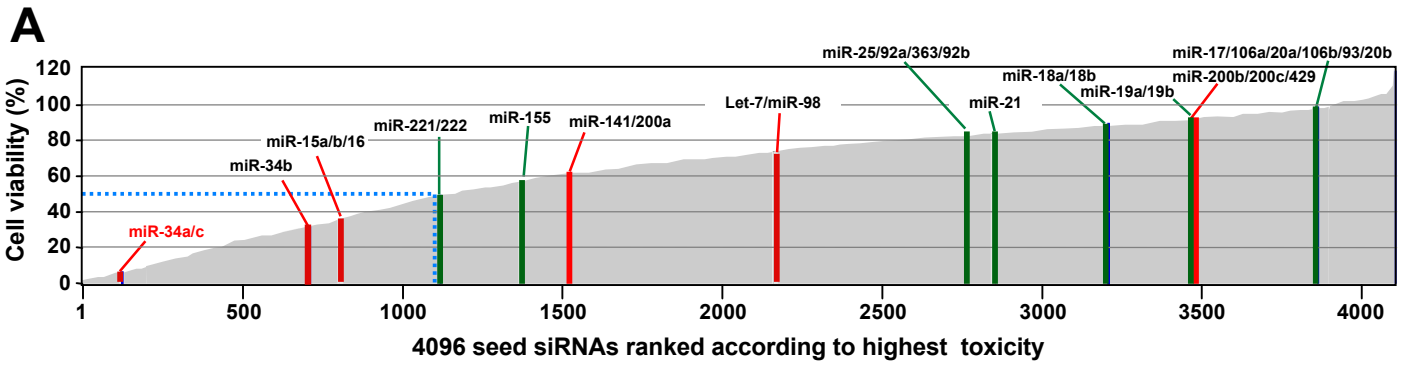
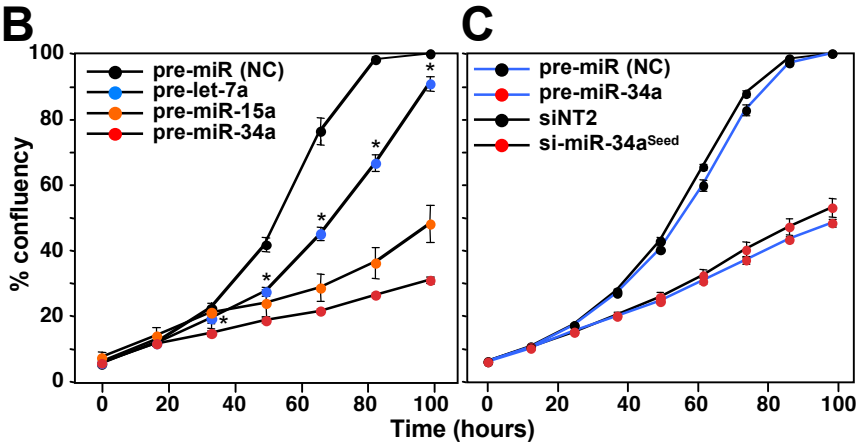
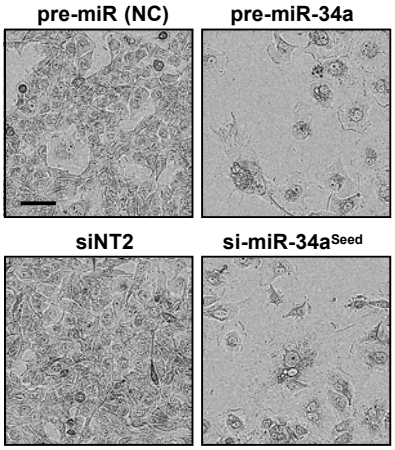
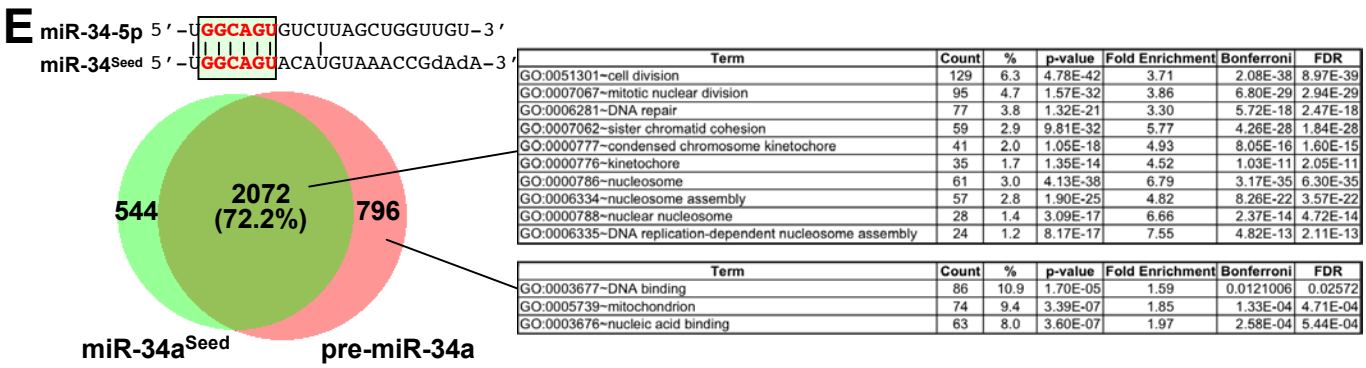


Figure 3



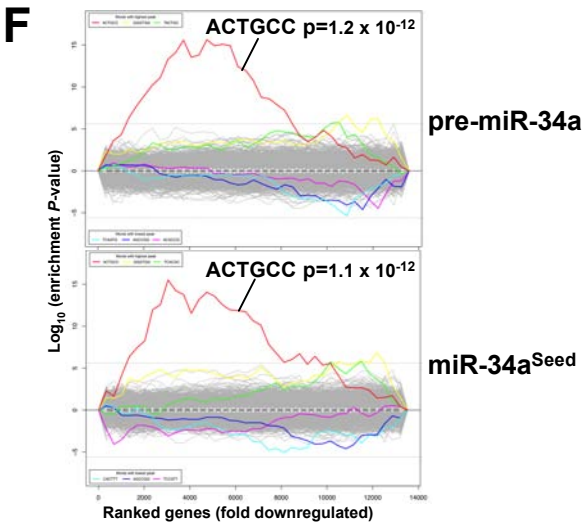






Term	Count	%	p-value	Fold Enrichment	Bonferroni	FDR
GO:0051301~cell division	129	6.3	4.78E-42	3.71	2.08E-38	8.97E-39
GO:0007067~mitotic nuclear division	95	4.7	1.57E-32	3.86	6.80E-29	2.94E-29
GO:0006281~DNA repair	77	3.8	1.32E-21	3.30	5.72E-18	2.47E-18
GO:0007062~sister chromatid cohesion	59	2.9	9.81E-32	5.77	4.26E-28	1.84E-28
GO:0000777~condensed chromosome kinetochore	41	2.0	1.05E-18	4.93	8.05E-16	1.60E-15
GO:0000776~kinetochore	35	1.7	1.35E-14	4.52	1.03E-11	2.05E-11
GO:0000786~nucleosome	61	3.0	4.13E-38	6.79	3.17E-35	6.30E-35
GO:0006334~nucleosome assembly	57	2.8	1.90E-25	4.82	8.26E-22	3.57E-22
GO:0000788~nuclear nucleosome	28	1.4	3.09E-17	6.66	2.37E-14	4.72E-14
GO:0006335~DNA replication-dependent nucleosome assembly	24	1.2	8.17E-17	7.55	4.82E-13	2.11E-13

Term	Count	%	p-value	Fold Enrichment	Bonferroni	FDR
GO:0003677~DNA binding	86	10.9	1.70E-05	1.59	0.0121006	0.02572
GO:0005739~mitochondrion	74	9.4	3.39E-07	1.85	1.33E-04	4.71E-04
GO:0003676~nucleic acid binding	63	8.0	3.60E-07	1.97	2.58E-04	5.44E-04



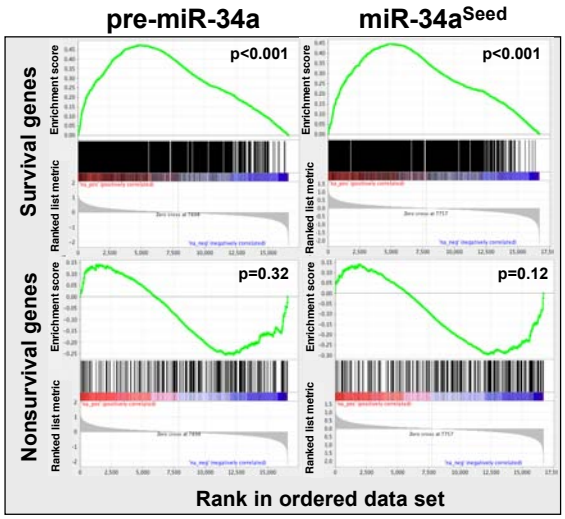
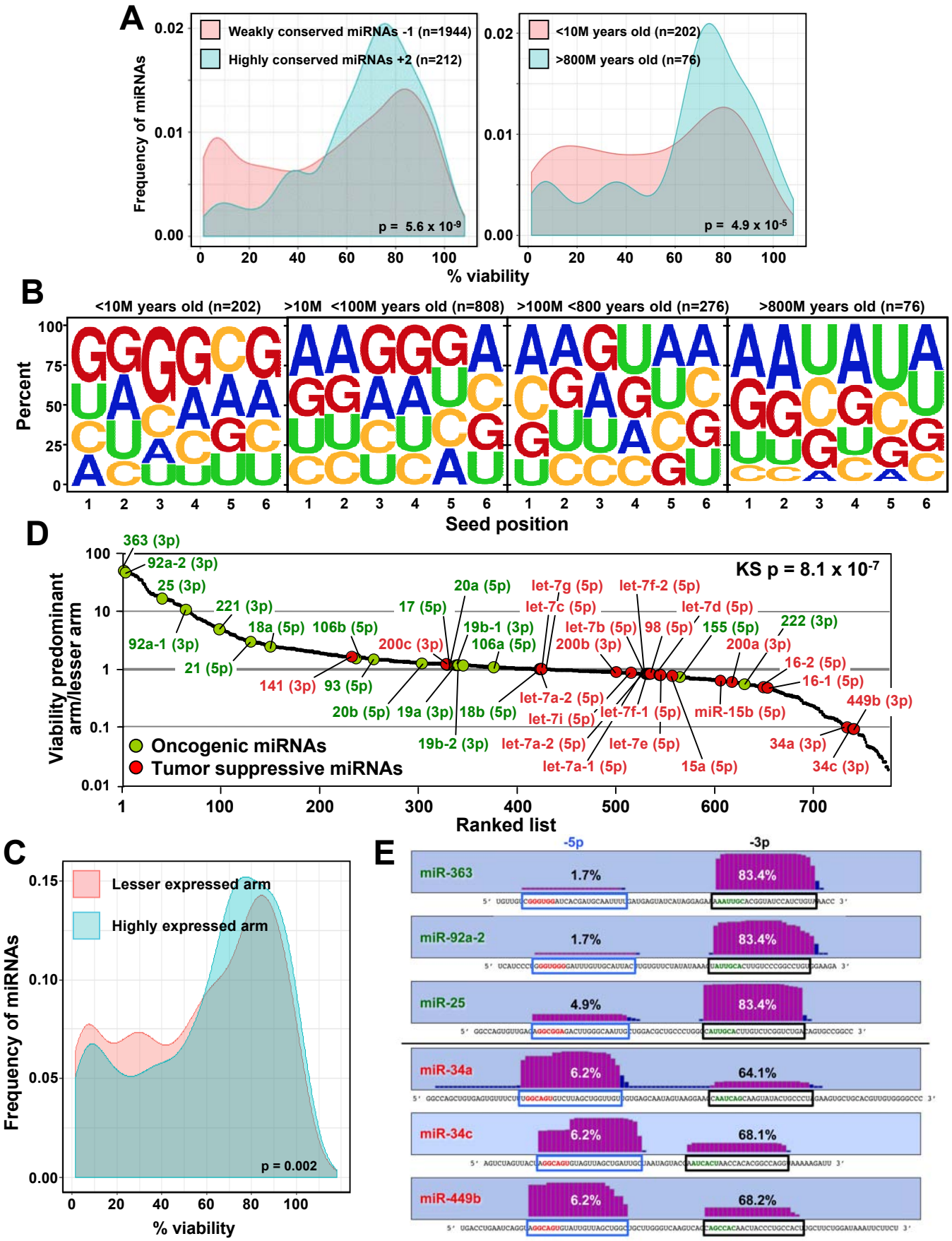


Figure 4


A
Weakly conserved miRNAs -1 (n=1944)
Highly conserved miRNAs +2 (n=212)
p = 5.6 x 10-9
<10M years old (n=202)
>800M years old (n=76)
p = 4.9 x 10-5
Frequency of miRNAs
% viability
B
<10M years old (n=202)
>10M <100M years old (n=808)
>100M <800 years old (n=276)
>800M years old (n=76)
Percent
Seed position
D
KS p = 8.1 x 10-7
Viability predominant arm/lesser arm
Ranked list
Oncogenic miRNAs
Tumor suppressive miRNAs
363 (3p)
92a-2 (3p)
25 (3p)
92a-1 (3p)
221 (3p)
21 (5p)
18a (5p)
106b (5p)
141 (3p)
93 (5p)
200c (3p)
20b (5p)
19a (3p)
17 (5p)
19b-2 (3p)
18b (5p)
20a (5p)
19b-1 (3p)
106a (5p)
let-7g (5p)
let-7c (5p)
let-7a-2 (5p)
let-7i (5p)
let-7a-2 (5p)
let-7a-1 (5p)
let-7f-2 (5p)
let-7b (5p)
200b (3p)
let-7f-1 (5p)
let-7e (5p)
98 (5p)
15a (5p)
let-7d (5p)
155 (5p)
miR-15b (5p)
222 (3p)
200a (3p)
16-2 (5p)
16-1 (5p)
449b (3p)
34a (3p)
34c (3p)
C
Lesser expressed arm
Highly expressed arm
p = 0.002
Frequency of miRNAs
% viability
E
-5p
-3p
miR-363
1.7%
83.4%
miR-92a-2
1.7%
83.4%
miR-25
4.9%
83.4%
miR-34a
6.2%
64.1%
miR-34c
6.2%
68.1%
miR-449b
6.2%
68.2%

Supplemental data

6mer Seed Toxicity Determines Strand Selection in miRNAs

Quan Q. Gao, William E. Putzbach, Andrea E. Murmann, Siquan Chen, Giovanna Ambrosini, Johannes M. Peter, Elizabeth T. Bartom, and Marcus E. Peter

Supplemental Figures

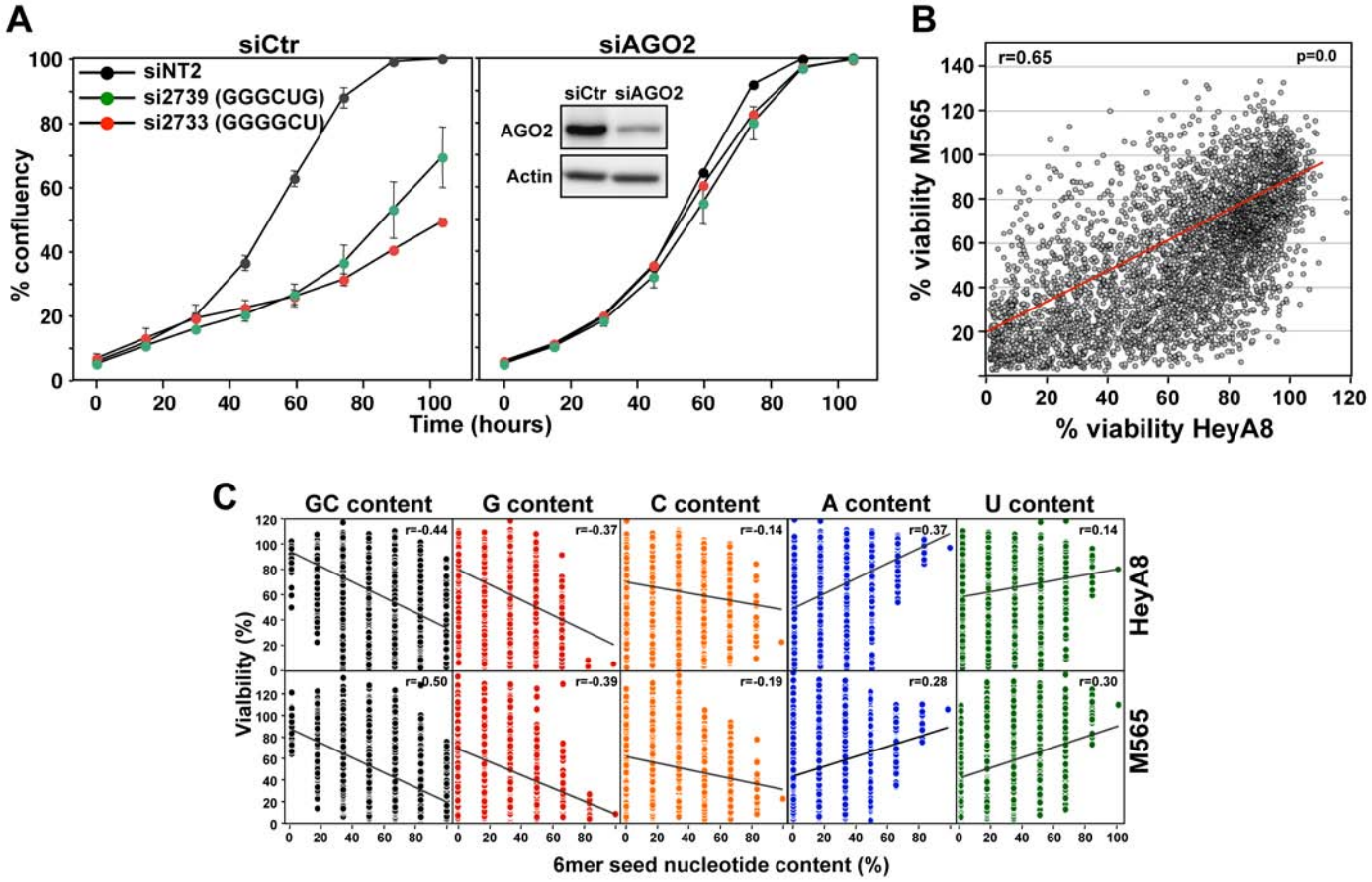


Figure S1: Toxic 6mer seeds prefer G-rich nucleotide composition.

(**A**) Percent cell confluence over time of HeyA8 parental cells transfected with either nontargeting control (siNT2) or the two most toxic siRNA duplexes (si2739 and si2733) at 10 nM. Cells were pretreated with either nontargeting SMARTpool siRNAs (siCtr, left) or AGO2 SMARTpool siRNAs for 24 hours at 25 nM (right). Insert shows Western blot to document AGO2 knockdown efficiency.

(**B**) Regression analysis showing the correlation between the 6mer seed toxicity observed in the mouse liver cancer cell line M565 (y axis) and the matching 6mer toxicity observed in the human ovarian cancer cell line HeyA8 (x axis). Toxicity of siRNAs and miRNAs in both cell lines can be interrogated at http://www.6merdb.org.

(**C**) Regression analysis of cell viability of 6mer seeds in HeyA8 cells (top panel) or M565 cells (bottom panel) versus GC content or individual nucleotide content (G, C, A, U) of the seeds. All p-values of the Pearson correlation coefficients were 0.0.

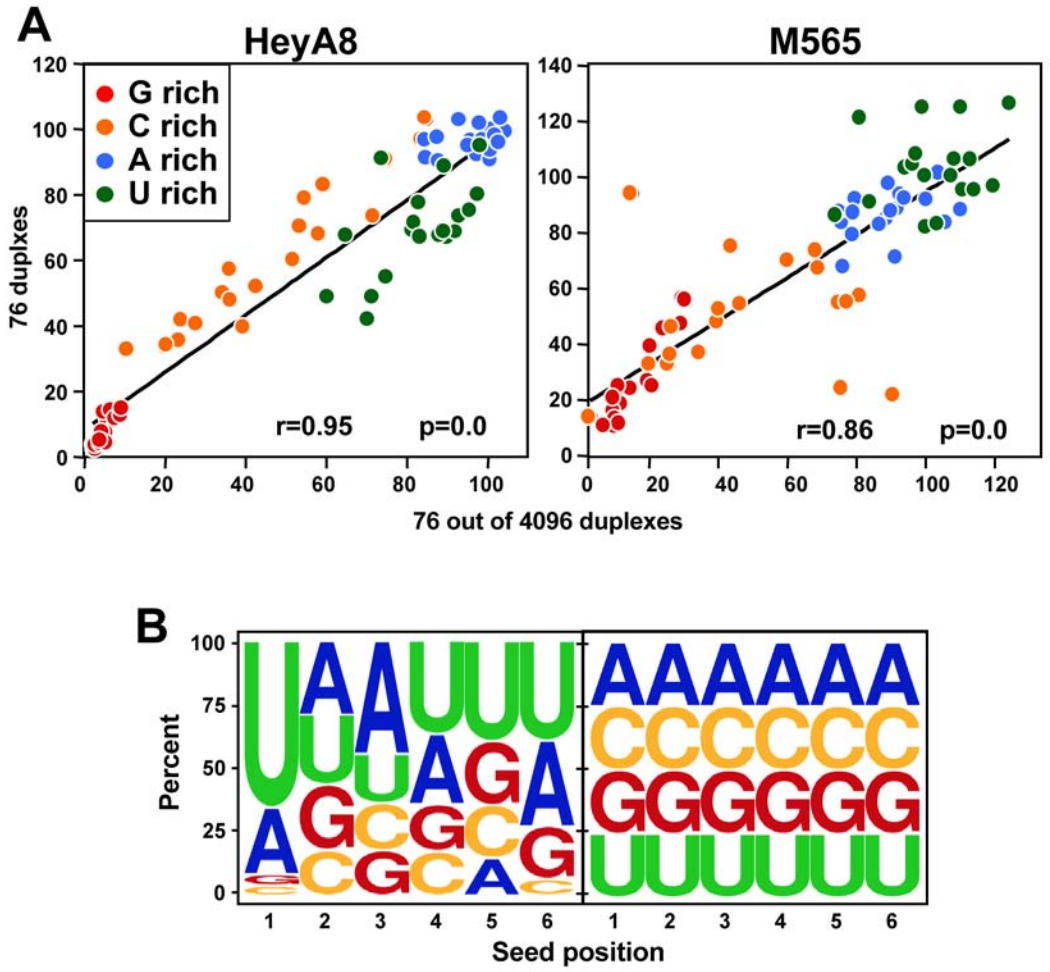


Figure S2: Reproducibility of the siRNA screens.
(**A**) Regression analysis showing the correlation in cell viability between the original 4096 duplexes screen (x-axis) and the repeat screen with the 76 duplexes with the highest nucleotide content in the seed region in either HeyA8 cells (left panel) or M565 cells (right panel).
(**B**) Nucleotide composition at each of the 6 seed positions in a commercial human genome-wide siRNA library (>65,000 siRNAs, left panel) or our balanced 6mer seed duplexes library (right panel).

A

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0050907	detection of chemical stimulus involved in sensory perception	1.22E-105	1.85E-101	7.82
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	1.18E-100	8.98E-97	8.42
GO:0009593	detection of chemical stimulus	1.11E-95	5.60E-92	6.9
GO:0050906	detection of stimulus involved in sensory perception	7.82E-90	2.96E-86	6.49
GO:0051606	detection of stimulus	2.07E-68	6.28E-65	5.23
GO:0007186	G-protein coupled receptor signaling pathway	7.50E-49	1.89E-45	4.49
GO:0007608	sensory perception of smell	3.19E-22	6.91E-19	12.7
GO:0007606	sensory perception of chemical stimulus	1.10E-17	2.09E-14	7.08
GO:0006959	humoral immune response	4.51E-15	7.59E-12	3.2
GO:0042742	defense response to bacterium	1.07E-14	1.62E-11	2.99
GO:0009617	response to bacterium	9.59E-14	1.32E-10	2.81

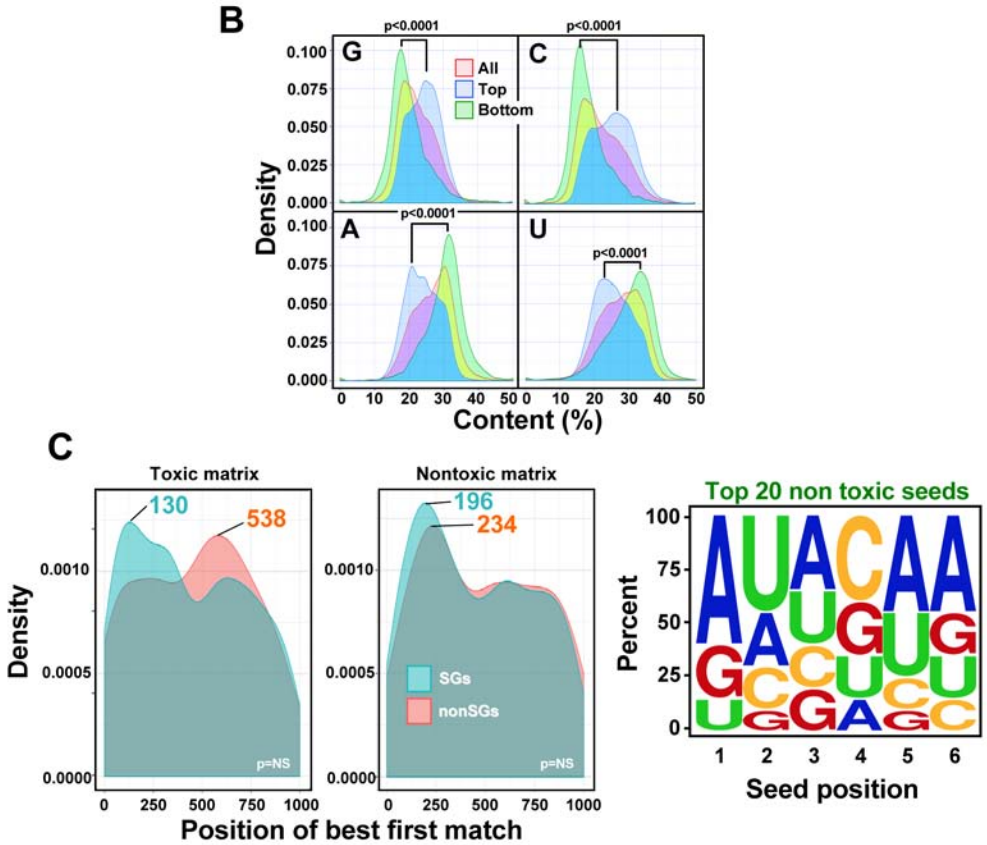
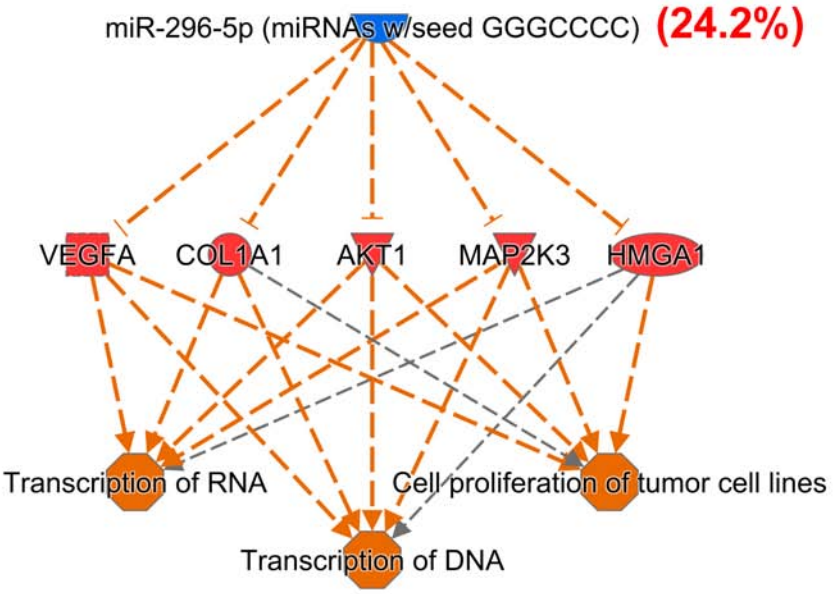


Figure S3: Link between high toxic PWM score and nucleotide content in survival and nonsurvival genes. (**A**) Results of a GOrilla gene ontology analysis using a gene list ranked by the PWM matrix match score with the toxic matrix in their 3'UTRs ranked from the lowest to the highest score. Only GO clusters that had a significance of enrichment of at least 10^{-11} are shown.

(**B**) Comparison of nucleotide content in 3' UTRs of genes with either a high toxic matrix match score (>400, n=4288) (Top), or a low matrix match score (<90, n=4834) (Bottom). Density plot of individual nucleotide content of all genes is shown for comparison. K-S test was used to calculate p-values for the difference between top and bottom genes.

(**C**) Distribution of the location of the best first match in target sites of either the toxic matrix (left panel) or the nontoxic 6mer seed matrix (right panel) in the 3' UTRs of survival genes (SGs) or nonsurvival genes (nonSGs). The analysis was performed as described in **Figure 2D**, right panel. The nontoxic matrix is shown on the far right. Peak maxima are given. NS= not significant.



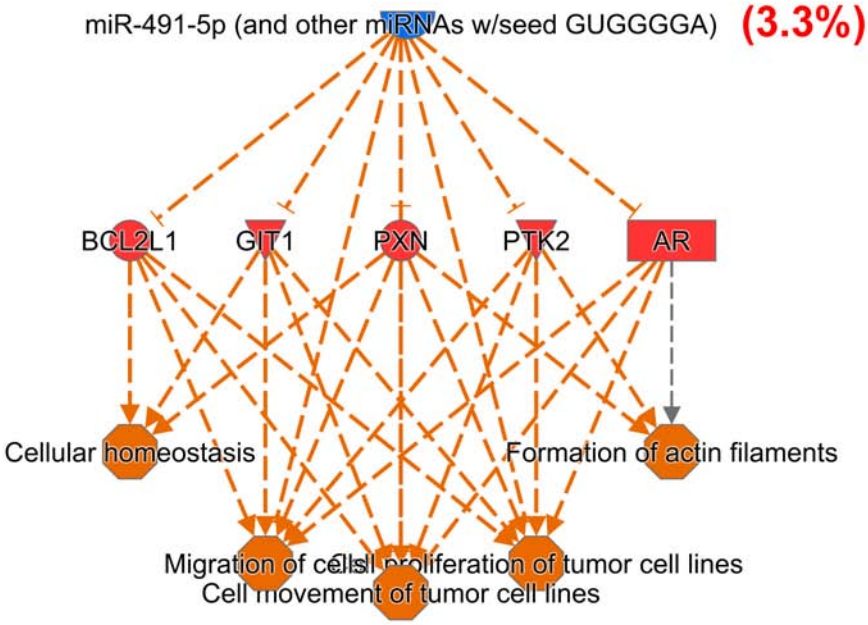


Figure S4: Ingenuity pathway analysis identifies miRNAs with G rich seeds to target survival genes. Top two Top Regulator Effect Networks identified by IPA regulated by miRNAs with G-rich seeds. The top 4288 genes with the highest PWM score (>400) were analyzed. Seed toxicity (percent survival) of the 6mer of the shown miRNA is given in red.

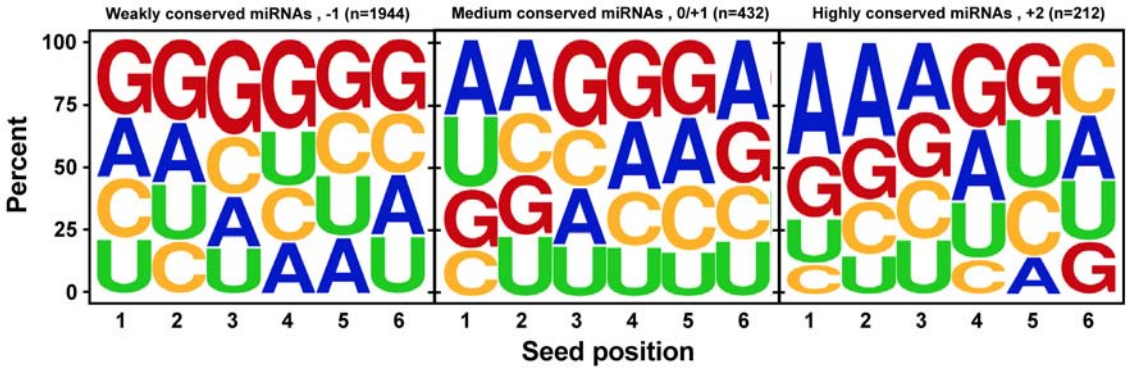


Figure S5: Loss of Gs with increasing conservation of miRNAs.
Nucleotide composition of each of the 6 seed positions in either poorly conserved (left), moderately conserved (center), or highly conserved miRNA seed families (right).
Note: The most toxic CD95L derived siRNAs we tested, siL3 (UAUGGG) and siL2 (GAGAGC), and miR-34a-5p (GGCAGU), all contain three Gs in their 6mer seed.

Supplemental Tables

Table S1: Results of the seed toxicity screens in HeyA8 and M565 cells. The top 200 most toxic and bottom 200 least toxic duplexes (to HeyA8 cells) are highlighted in red and green, respectively. An additional tab lists the siRNA duplexes that were super toxic to both HeyA8 and M565 cells.

Table S2: Matrixes used for the PWM analysis. Nucleotide compositions of the 20 most toxic and 20 least toxic 6mer seed duplexes that was used to generate the PWM matrices (the reverse complement of these matrices).

Table S3: Results of RNA Seq analysis of HeyA8 cells transfected with either pre-miR-34a or miR-34a^{Seed}.

Table S4: Ratio of 6mer seed toxicity of the predominantly expressed versus the lesser expressed arm of all miRNAs.