מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

עבודת גמר (תזה) לתואר
מוסמך למדעים

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

מאת
יונית הופמן

Thesis for the degree
Master of Science

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

By
Yonit Hoffman

רצפי אַלוּ ומיקרורנ"א: יחסי הגומלין בגנום וברגולציה של מסלול העברת האותות p53

Alu sequences and microRNAs: Global interplay in the genome and in the regulation of the p53 network

מנחים:
צחי פלפל
ומשה אורן

Advisors:
Tzachi Pilpel and
Moshe Oren

תשרי תשע"א

October 2011

Table of Contents

Abstract

The massive spread of repetitive elements in the human genome poses a substantial challenge to the organism since they may accidentally contain seemingly functional elements placed outside their natural context. Nonetheless, such elements may also offer an opportunity to the species as they introduce potential new genetic material that could be used by further evolutionary refinement. A striking example is provided by the million copies of Alu repeats in the genome, of which ~0.5% reside within 3'UTRs of genes. Some of these elements may serve as raw material for evolutionary novelty, yet the cell needs to mask the rest from a potential distorting regulatory effect. In this work I took a genome-wide view of general regulatory effects of miR targets within Alus, and combined it with more targeted investigation of a specific network of the tumor suppressor p53 pathway, to investigate whether miR targets inside Alus insert new regulatory mechanisms. Focusing on Alu elements that contain seemingly canonical sites for a diversity of microRNAs I show, using a comprehensive dataset of microRNA functional assays, that the microRNA machinery appears to often ignore Alu-contained sites. I identified three potential means that may account for such apparent "repeat-masking": (i) functional miR sites and Alu repeats appear to have distinct mutually-exclusive territories within 3' UTRs, (ii) tight secondary structure of Alu-contained sites that limits access to the microRNA machinery, and (iii) deamination during RNA editing of Alu-contained microRNA sites. The combination of these three means appears to allow cells to ignore most Alu-contained potential microRNA sites and perhaps deal only with a selection of sites that may nonetheless become functional.

The p53 pathway is pivotal in tumor suppression. p53 leads cells to growth arrest, senescence and apoptosis via modulating the expression of various target genes. Two major regulators of p53 are MDM2 and MDM4. Both proteins inhibit p53 by direct binding, while MDM2 is also an E3 ubiquitin ligase that sends p53 to degradation by the proteasome. The delicate interplay between the levels of MDM2, MDM4 and p53 is very important for cell cycle control and cellular homeostasis. I describe two microRNAs that regulate MDM2 and MDM4: miR-339-5p and miR-661. miR-339-5p regulates MDM2 at the mRNA and protein levels, while miR-661 regulates MDM2 at the protein and mRNA levels and MDM4 at the protein level only. The targets of miR-661 in *MDM2* and *MDM4* are predominantly within Alus, presenting an evolutionary novel regulatory mechanism which is primates specific, and an example of the small fraction of miR targets within Alus that are functional.

The regulation of MDM2 and MDM4 by those miRs causes an increase in the functionality of p53, which leads to upregulation of p53 target gene expression and to cell cycle alterations.

Introduction

microRNAs

microRNAs (miRs) are small non coding RNAs, ~22 nt long[1], which regulate gene expression through interaction with their mRNA targets. There are hundreds of miRs in the human genome, for which thousands of predicted target sequences exist in virtually all encoded mRNAs. miRs in general, and specifically the ones that are conserved among mammals, are known to participate in many important cellular and developmental processes, as well as in the development of some diseases[2-3].

The processing of miRs involves different factors in different parts of the cell. The long miR primary transcript (pri-miRNA) is cleaved to a pre-miRNA of 60-70 nt by the nuclear RNAse III Drosha[1]. Drosha cleaves near the base of the primary stem-loop, in both strands of the stem[1].Then the pre-miRNA is exported to the cytoplasm by Ran-GTP and the export receptor Exportin-5, where it is cleaved by Dicer to produce the mature miR[1]. Dicer is an RNAse III endonuclease[1]. Dicer recognizes the double stranded part of the pre-miRNA with an affinity to the 5' phosphate and the 3' overhang at the base of the stem loop, and it cuts the strands of the duplex two helical turns from the base of the stem-loop[1]. The mature miR is then loaded onto the RISC complex to direct the complex to a specific subset of mRNA, thus inhibiting their translation or targeting them for cleavage and degradation[1].

It is generally accepted that most mRNAs recognized by a particular miR have a target in their 3'UTR that is complementary to bases 2-8 (called seed) in the mature miR 5' end[4]. Although each miR has potentially hundreds of target sites in at least dozens of mRNAs, it is clear that only a minority of those sites are functional, while the majority are non functional[5]. In addition, the functionality of the target may be influenced by the identity of the cell and the physiological conditions[6]. There exist numerous target prediction algorithms, which use different parameters such as conservation, structure, length of the target etc[7]. Still, it is not clear how the cell and specifically the miR machinery distinguish between the plethoras of potential targets and identify the genuine ones.

Alu sequences

Alu repeats ("Alus") are transposable elements with more than one million copies in the human genome [8]. The structure of the Alus is a dimer formed by the fusion of two monomers derived from the 7SL RNA gene[9]. Most Alus were inserted ~40 million years ago in the primate lineage[10]. There is an Alu insertion every 20 births[11]. Alu sequences are retrotransposons; they use the retrotransposition molecular machinery of the LINE TE- L1s to integrate into the host genome[12].

Seventy five percent of human genes contain an Alu sequence, most of them in introns[13], and 5% of the 3'UTR, on average, is comprised of Alus[14]. The Alu density in genes is higher than in intergenic regions except for chromosomes 19 and 22[13]. Transcripts of Alu are present at low level under normal conditions, but their levels increase when cells are exposed to a variety of stresses[15].

Recent studies show that Alus are not "junk DNA" as thought before. Over the years examples have accumulated that demonstrate how genomes converted Alus into a diversity of functional elements. Vansant et al. found that the consensus sequences of Alu classes III and IV, which are sometimes present within gene promoters, contain a transcription regulatory site (Retinoic Acid Response Element, RARE)[16]. During primate evolution, many Alus that were likely to have been inserted near genes where a RARE site was deleterious, were selected against[16]. Yet, some Alus were inserted upstream to genes for which Retinoic Acid inducibility was an advantage, and those stayed and are recognizable today, such as in the K18 gene[16]. Polac et al. found promoter regions that are enriched with Alus that contain multiple binding sites for transcription factors, many of which are associated with early markers of development[17]. They suggested that evolution used the Alus in order to insert regulatory elements (TF binding sites) into promoters[17]. Sorek et al. showed that mutations in the Alus can create new weak splice sites, so that the Alu is inserted into the protein in a process termed "Alu exonization"[18]. Examples include a transcript variant of the biliary glycoprotein that contains an Alu fragment, and the human decay-acceleration factor where 10% of its mRNA molecules contain an Alu[19].

Several functions were found for Alus in the 3'UTR. An et al. found that Alus in the 3'UTR create AU rich elements – AREs; nearly 40% of the AREs are associated with Alus[14]. AREs destabilize the mRNA through the nuclear exosome pathway[20,21]. Gong et al. showed that imperfect binding of an Alu element in the 3' UTR of a SMD target (a target for Staufen 1 (STAU1)-mediated messenger RNA decay, which mediates decay) and another Alu element in a cytoplasmic, polyadenylated long non-coding RNA (lncRNA) can form a binding site for STAU1, and therefore affect the mRNA

stability[22]. Chen et al. found that Alus contain specific hexamers that are polyadenylation sites (PAS) directly or with one base mutation[23]. There are about 107 Alus that gave rise to such active PAS[23].

The connection between miRs and Alus starts in the birth of some miRs. Several reports describe the possible birth of novel miRs from genomic DNA repeats, and this has been estimated to account for 10-20% of the miRs[24,25].

In the C19MC Alus-miRs cluster, there are many miRs surrounded by Alu sequences, embedded in cassettes that are repeated[26]. Alu sequences contain an internal promoter for RNA polymerase III[27], and Borchert et al. showed that Pol III transcribes the Alus in the cluster, and as there is no termination sign it also transcribes the miRs. Three years later, Bortolin-Cavaillé et al challenged this claim and showed that Pol II is responsible for transcribing the miRs in this cluster. Lehnert et al. proposed a model in which the homology sites of Alus mediated the duplication of the cassettes that comprise the miRs in this cluster. Smalheiser and Torvik described several mammalian miRs that are derived from genomic repeats, and as such have target sequences within these repeats[28]. Specifically, they identified many mRNAs that contain Alus in their 3'UTR, within which there are targets for dozens of miRs[29]. Lehnert et al. found that there is a small number of miRs with more than 1000 targets per megabase in Alu sequences. 10 of the 17 miRs that are enriched with targets within Alus are from the C19MC cluster[30]. Their hypothesis was that the miRs with the enrichment of targets in Alus protect against Alu transposition[30]. This hypothesis was based on the observation that miRs of the C19MC cluster are expressed significantly in the testis[30]. Furthermore, Shalgi et al suggested that miRs have a role in maintaining genomic stability by the repression of transposable elements[31].

As Alus are very abundant in the human genome, the potential effect of miR targets within them may be substantial and affect crucial cellular processes.

p53, MDM2 and MDM4

p53 is a transcription factor which responds to diverse stresses such as DNA damage, overexpressed oncogenes and various metabolic limitations[32-33]. p53 regulates the expression of a diverse group of genes that initiate cell cycle arrest, senescence, apoptosis, metabolism alteration and DNA repair[33]. By regulating these genes, p53 prevents the proliferation of genetically compromised cells[34]. The p53 pathway is very important for tumor suppression in humans. Notably, p53 is mutated in ~50% of human cancers, and functionally inactivated in many more[35].

Senescence can be triggered by oncogene activity or DNA damage. Many oncogenes such as RAS, E2F, RUNX1 and more trigger p53-induced senescence[36]. Some of them involve the DNA damage response (DDR), while others induce p53-dependent senescence without DNA damage[36]. DNA damage such as radiation, chemotherapeutic drugs or telomerase dysfunction drive senescence primarily via the p53-p21 pathway[36]. p21 is important for DNA damage induced senescence as well as for transient growth arrest[36]. Biochemically, p21 is a cyclin dependent kinase inhibitor, whose induction causes a G1 arrest. This G1 arrest is critical for genomic integrity, since it prevents the cell from replicating damaged DNA[37]. Specifically, p21 inhibits cdk2 and cdk4, which are required for the progression of the cell cycle from G1 to S phase[38].

p53 controls cell cycle progression also by regulating the G2/M checkpoint, which is very important for preventing segregation of damaged chromosomes[38]. This checkpoint can be inhibited by controlling Cdc2-cyclinB activity[38]. p21 also binds to the Cdc2-cyclinB complex and prevents its activity. Furthermore, p53 can also regulate additional target genes that control this checkpoint, such as 14-3-3 sigma , GADD45, BTG2, REPRIMO, B99 (*GTSE-1*), and more[38].

The mechanisms that are responsible for the activation of p53 in senescent cells are not completely understood. However, some candidates are emerging. One such mechanism is an increase in the expression of the tumor suppressor ARF, which binds to MDM2 and inhibits its activity, thereby preventing p53 degradation[39]. Another activator is the tumor suppressor PML (promyelocytic leukemia); expression of PML is also regulated by p53, creating a positive feedback loop between them[40]. PML stabilizes p53 through increasing its acetylation[41-42].

DNA damage repair is a very important process, since failure to repair damaged DNA results in cell death or oncogenic transformation. p53 regulates the DNA repair process through p53 target genes that participate in this process, as well as by modulating the process directly[38].

Apoptosis has a key role in tumor suppression. p53 triggers apoptosis in response to severe stress, and this involves transcriptional activation of proapoptotic genes[36]. Such p53 targets genes encode proteins that participate in both the intrinsic and the extrinsic apoptotic pathways[36]. Among the intrinsic pathway genes are Bax, Bid, Puma and Noxa. Among the extrinsic pathway genes are the Fas (CD95) and DR5 death receptors, the death ligand TNFSF10, the Fas ligand TNFSF6 and caspase 8[36]. p53 can also promote apoptosis by direct apoptogenic effects on the mitochondria [38].

MDM2 and MDM4 are structurally related proteins[43]. They both contain an amino-terminal domain that binds to p53, as well as a central acidic domain and a carboxy-terminal RING finger through

which they form heterodimers[44]. In general, MDM2 loss is almost invariably lethal, while the loss of MDM4 in some cases is tolerated[35]. MDM2 and MDM4 are overexpressed in different tumors[45]. These differences reflect the different roles of MDM2 and MDM4 in tumorigenesis and stress response[43].

MDM2 is part of a family of ring finger containing proteins[34]. It is an E3 ubiquitin ligase, whose catalytic activity requires its ring finger for both autoubiquitylation and p53 ubiquitylation[46]. MDM2 binds to p53 and inhibits p53-mediated transactivation[47] both by physically masking the transactivation domain of p53 and by promoting ubiquitylation and proteasome-dependent degradation of p53[48]. The masking of the transactivation domain is achieved through the binding of MDM2 to an N-terminal region of p53, which blocks p53 from associating with the transcriptional machinery [49]. The *MDM2* gene contains two p53 binding sites, which allow p53 itself to control the levels of *MDM2* mRNA[50].

The MDM2-MDM4 hetero-oligomer is a more efficient E3 ligase of p53 than MDM2 alone, meaning that MDM4 acts as an enhancer of the E3 activity of MDM2 towards p53[51]. MDM4 can either stimulate or inhibit MDM2 E3 ligase activity, depending on the circumstances[52]. MDM4, unlike MDM2, does not contain a p53 binding element in its gene, and therefore *MDM4* mRNA levels are not regulated by p53[53].

MDM4 can bind to the p53 transactivation domain, limiting the access of essential transcriptional coactivators and of the transcription machinery[54]. For example, this binding prevents the interaction of p53 with the histone acetyl transferase p300 and reduces the acetylation of p53, which is involved in its activation [55]. Unlike MDM2, it does not possess E3 ubiquitin ligase function[56].

Upon DNA damage, the MDM2 ubiquitin ligase activity is directed preferentially towards MDM4, leading to degradation of the endogenous MDM4 protein and promoting p53 activation[57]. Another mode of p53 activation takes place upon nucleolar disruption. Under such circumstances, ribosomal proteins are released. Some of those proteins interact with MDM2 (but not with MDM4), and inhibit its ubiquitylation activity towards p53[57]. Upon such stress MDM4 is also degraded by MDM2, albeit in a different manner than upon DNA damage[57].

Transcription of the *MDM4* gene is stimulated by mitogenic signaling, particularly through the MEK/ERK pathway. Likewise, expression of oncogenic K-RAS also causes activation of the *MDM4* promoter[57].

miR-661

miR-661 is a human miR (without an ortholog in mouse). There are three established targets of miR661 in human: MTA1, StarD10 and Nectin1.

MTA1 is upregulated in different tumors, such as gastrointestinal cancers, including esophageal, gastric and colorectal cancers[58]. Overexpression of MTA1 is associated with tumor angiogenesis and higher tumor grade, and it was suggested as a predictor of aggressive phenotype[59]. Reddy et al showed, by western blot and luciferase assay, that the levels of MTA1 protein were downregulated when miR-661 was overexpressed[60]. They further showed that the transcription factor c/EBPalpha binds to miR-661 to positively regulate its levels, which further downregulates the levels of MTA1[60]. The levels of MTA1 increase in a breast cancer progressive isogenic model as the tumors progress, while miR-661 levels decrease[60]. MTA1 levels were shown to correlate with those of the transactivator protein HBx[61], which is a regulator of cellular responses to hepatitis B virus. HBx is an activator of NF-κB signaling[62]. It also enhances the expression of inducible nitric-oxide synthase (iNOS)[63,64]. Bui-Nguyen et al showed that miR-661 is downregulated by HBx, to upregulate the levels of MTA1[61]. They further showed that the iNOS pathway is regulated by miR-661 indirectly, via its regulation of MTA1[65].

In contrast to the above data, which suggest that miR-661 is downregulated during tumor progression and is therefore behaving as a candidate tumor suppressor, Vetter et al showed that miR-661 contributes positively to the invasive behavior of breast cancer cells[66]. miR-661 targets StarD10, a lipid transferase, and Nectin1, a cell-cell adhesion protein[66]. The levels of miR661 are elevated in highly invasive breast cancer cell lines such as MDA-MB-435 and MDA-MB-231, while in weakly invasive and non invasive cell lines (such as MCF7 and MCF10F) it is expressed at lower levels[66]. Thus, miR-661 may either inhibit or promote cancer-related processes, depending on the cellular context.

miR-339

miR-339 is a human microRNA with an ortholog in mouse. There are a few established targets of miR-339 in the human transcriptome.

Ueda et al showed that Dicer is responsible for the generation of mature miR-339, which targets ICAM-1, a cell adhesion protein[67]. The effect of downregulation of ICAM-1 is that tumor cells are less susceptible to CTL-mediated cytolysis[67]. Ju et al found that miR-339 showed an upregulation of

3.4 fold in pre-B-ALL (childhood B-cell precursor acute lymphoblastic leukemia) compared to normal marrow[68]. They suggested that the overexpression of the miR is possibly linking the transcriptional pathogenesis of pre-B-ALL[68]. In contrast, Wu et al showed that the expression of miR-339-5p is low in the aggressive cell lines MDA-MB-468 and MDA-MB-231 and in breast cancer tissues[69]. They reported that overexpression of miR-339-5p reduced cancer cell migration and invasion capacities, while downregulated levels of the miR were associated with an increase in metastasis to lymph nodes and with high clinical stages[69]. BCL-6 has a target for miR-339-5p and was downregulated after overexpression of the miR[69]. Ichi et al found that miR-339-5p downregulates the histone demethylase enzyme KDM6B by 20-35% in ND7 cells (mouse neuronal cell line), while when in combination with miR-185 and miR-148a or miR-138 the downregulation was up to 90%[70].

Research goals

In my research project I wished to address the following questions:

- To examine whether miR sites within Alus are functional
- In cases where miR sites are non-functional discover potential mechanisms by which sites are masked
- Identify novel miR regulators of the p53 network
- Focus on a novel pair of miRs that reside within Alus in UTRs of the p53 network and examine their potential biological effect

Materials & Methods

Cell culture, plasmids, siRNA, miR transfections and DNA damage

Cells were maintained at 37°C. MCF7 cells were grown in DMEM (Biological Industries) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Hyclone) and penicillin-streptomycin antibiotics solution (all from Biological Industries).

miR and siRNA (20 nM, Dharmacon) transient transfection was done according to the manufacturer's instruction with Dharmafect1 reagent (Dharmacon).

Doxorubicin (Sigma) was added to the culture medium at a final concentration of 1μM. Etoposide (Sigma, in DMSO) was used at a final concentration of 50μM. Both these genotoxic agents were added 48h after miR transfection.

RNA purification, Real-Time quantitative PCR and Microarray hybridization

RNA was extracted with the mirVana miRNA Isolation Kit (Ambion), for Real-time reverse transcription (RT)-PCR analysis. From each RNA sample 1.5μg was reverse transcribed with Moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI), random hexamer primers (Sigmna) and dNTPs (LAROVA). Real-time PCR was done on a StepOne real-time PCR machine (Applied Biosystems, Foster City, CA) with Syber Green PCR supermix (Invitrogen).

Primers

Reverse primer	Forward primer	Gene
GGCATTCTGGGAGCTTCATCT	CCCAAGCAATGGATGATTTGA	p53
CACGATACCAAAGTTGTCATGGAT	AGCCTCAAGATCATCAGCAATG	GAPDH
ACTGCCACTCATCCTCAGAGGTA	AATGATGACCTGGAGGACTCTAAGTC	MDM4
GCGGATTAGGGCTTCCTCTT	GGCAGACCAGCATGACAGATT	p21
GGTTACAGCACCATCAGTAGGTACAG	CAGGCAAATGTGCAATACCAAC	MDM2
TTGATGTCAGTCACTTGGGCAT	CCCTCCTACCTCTGGTTCTTACG	CD95
GCCCTTGGACGGCTTTTC	CCAGGAGGCACTCACAGAGC	Btg2
TCTGCGGAGGGACTGGAAC	AGCTGTCCTCCTCCTGCTAGAA	Wig1

Antibodies

The following primary antibodies were used in western blot analysis:

GAPDH monoclonal antibody was purchased from Millipore (MAB374)

The monoclonal antibodies 4B2, 2A9, and 4B11 were used to detect MDM2.

BL1258 (Bethyl Laboratories) was used as anti-MDM4 antibody.

The monoclonal antibodies PAb1801 and DO1 were used to detect p53.

Anti-p21 antibody (c-19) was purchased from Santa Cruz Biotechnology.

Western blot analysis

For western blot analysis cells were washed with PBS, collected and lysed with NP40 lysis buffer (150mM Sodium Chloride, 50mM Tris pH=8, 1% NP40) with protease inhibitor mix (Sigma). Cells in suspension were vigorously vortexed and centrifuged at 14000 RPM for 10 minutes at 4°C, and the soluble fraction was used to determine protein concentration in each sample. The protein concentration was determined with a BCA kit (Thermo scientific) according to the manufacturer's protocol. Protein sample buffer (3% SDS, 10% glycerol, 5% β-mercaptoethanol, 62mM Tris pH=6.8) was added, and samples were boiled for 5 minutes and loaded onto SDS-polyacrylamide gels. Proteins were transferred onto nitrocellulose membranes, followed by 30 minutes blocking in 5% milk in PBS. The membranes were incubated with primary antibodies overnight at 4°C, washed 3 times with PBST (0.05% Tween-20 in PBS), and reacted for 45 minutes with horseradish-peroxidase (HRP)- conjugated IgG, followed by 3 washes with PBST and one wash in PBS. The proteins were visualized using an enhanced chemiluminescence (ECL) detection kit (Amersham).

Luciferase assay

Cells were seeded in 12-well dishes, and each well was transfected with with 200 ng of firefly luciferase reporter plasmid DNA (p21 WT or p21 mutated) and 40ng renilla luciferase plasmid DNA, using the JetPei reagent (Polyplus Transfection) in NaCl (Polyplus Transfection), according to the manufacturer's protocol. Cells were washed twice in PBS and lysed with passive lysis buffer (Promega, USA) for 15 minutes with shaking. Luciferase reporter activity was measured in a luminometer (Moduluc Microplate, Turner BioSystems).

FACS cell cycle analysis

Cells were detached by trypsinization, harvested in fresh medium, and combined with the removed culture medium that included the floating cells. Cells were pelleted, washed with cold PBS and fixed with cold 70% ethanol for at least 30 minutes at -20ºC. Ethanol was removed and cells were then washed with PBS twice, and resuspended in 100µl PBS containing 50µg/ml 50µg/ml RNase A for 15 minutes. Then 400µl PBS with propidium iodide (PI, Sigma, 50µg/ml) was added and cells were kept in the dark for at least 10 minutes. Cells were analyzed by flow cytometry (PI: excitation- 488nm, emission- 585nm) for DNA content, using a FACS instrument (FACSort, Becton Dickinson).

Predicting miRNA targets in the p53 network

The list of p53 network genes was taken from KEGG (http://www.genome.jp/kegg/) with some additions. All 7-11-kmers were scanned using JAVA code in the 3'UTR to find candidate motifs that may serve as miRNA targets. The binomial test was calculated using Java Statistical Classes (http://www.jsc.nildram.co.uk/index.htm). FDR was calculated using MATLAB.

Human and mouse genomes and Alu sequences

The full human 3'UTR sequence data was taken from UCSC[71], NCBI36/hg18. Alu sequences and their locations were taken from Repeat Masker (only Alu family) [90]. The mouse 3'UTRs sequences and coordinates were taken from UCSC[71] (NCBI37/mm9). The mouse repeats were taken from Repeat Masker[90].

Prediction of miR targets and conservation analysis

miR target sites were predicted by scanning for the seed of the miR, on the basis of perfect Watson-Crick complementary. Targets were defined as perfect 7-mer, for all human and mouse miRs listed in miRBase release 15 (http://www.mirbase.org/). Conserved targets were taken from TargetScan release 5.1 (http://www.targetscan.org/), m8 target type only (perfect 7-mer). In the analysis of conservation and folding energy the only genes that were included were genes of the following types: (1) the 3'UTR in the UCSC version was the same as the 3'UTR of the genes used by TargetScan (as defined in their website); (2) The 3'UTR of the genes from UCSC were included inside the 3'UTR defined in TargetScan or the opposite.

Analysis of the Khan data

The data regarding the miR-overexpression experiments was taken from Khan et al, 2009. This data contains siRNAs as well as miR overexpression experiments. Here we analyzed a subset of 43 miR over-expression experiments. For each miR that was over-expressed, site was scanned against all human 3'UTR. Percent down-regulation of target mRNAs was defined as the percentage of genes with fold reduction of at least 1.6 (i.e. 0.7 of a log2 scale), except for figure 16 where a cut-off of 1 was used.

Secondary structure prediction

Secondary structure were predicted for all human and mouse 3'UTRs using the Bioinformatics Toolbox of Matlab 10. The analysis was done in windows of 100 bps, and up to 50 bps from the last coding exon were added to the beginning of the 3'UTR for the prediction. The structure of each nucleotide of the 3'UTR was determined according to its structure in the folding where this nucleotide was in position 51 (in the window of 100). The folding energy of each nucleotide was the average folding energy of this nucleotide in all the folding windows in which it was included.

Results

Analysis of a dataset of genome-wide expression array experiments after miRs overexpression

The mechanism by which miRs regulate mRNA function are under extensive investigation, as are the parameters that determine productive miR-mRNA recognition. Khan et al. have recently assembled the results of a series of genome-wide expression array experiments into a single normalized database[72]. I analyzed this database in order to obtain new insights into the regulation of specific mRNAs by miRs.

For all the following analyses I used a subset of experiments from the Khan database, comprising 43 experiments with 23 different miRs overexpressed in a total of five different cell lines. In each experiment the dataset provides information on the genome-wide mRNA response to the overexpression of one miR at a time in a given cell line.

In order to set the fold change threshold that defines a downregulated gene I first looked at the fold change average of each experiment for all the genes with fold change lower than 0 (log2 scale) in two groups of genes: genes that contain the miR binding site, and genes without the miR binding site. As can be seen in figure 1, a threshold of -0.7 (log2 fold change) discards the off-target effects of the miR, since in all of the experiments the average fold change for genes without the miR binding site is above -0.7.

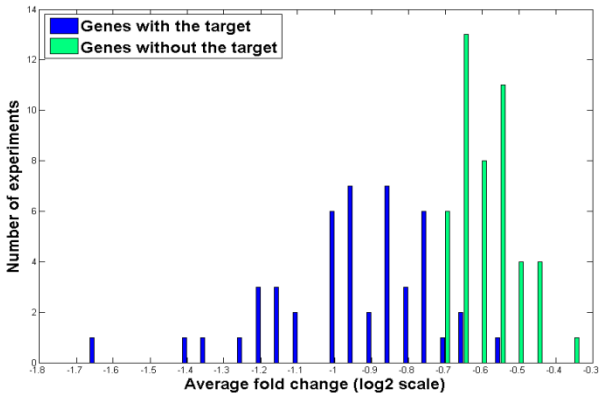

Figure 1: Average downregulation of genes in the microarray experiments. The average fold change (log2 scale) in the levels of the corresponding transcripts in each experiment was calculated for genes with fold change lower than 0 (log2 scale) separated to two groups of genes: those that contain the miR target and those that lack it.

miR targets in the 3'UTR are important for the functionality of the miR

In order to further establish the intuitive notion that the presence of a miR target is required for downregulation by that miR, I first examined whether genes that contain a miR binding site are more likely than other genes to be downregulated in response overexpression of that miR. Figure 2 shows that the percentage of downregulated genes was significantly higher in the group of genes that contained a sequence motif predicted to be bound by the overexpressed miR, in comparison to the group of genes without such binding site (Student's t-test P-value of 3.6e-16). In addition, the average fold change of the donwregulated genes was significantly higher in the group of genes containing the binding site compared to the downregulation that is occasionally observed among the control-set genes (Student's t-test P-value of 4.7e-10).

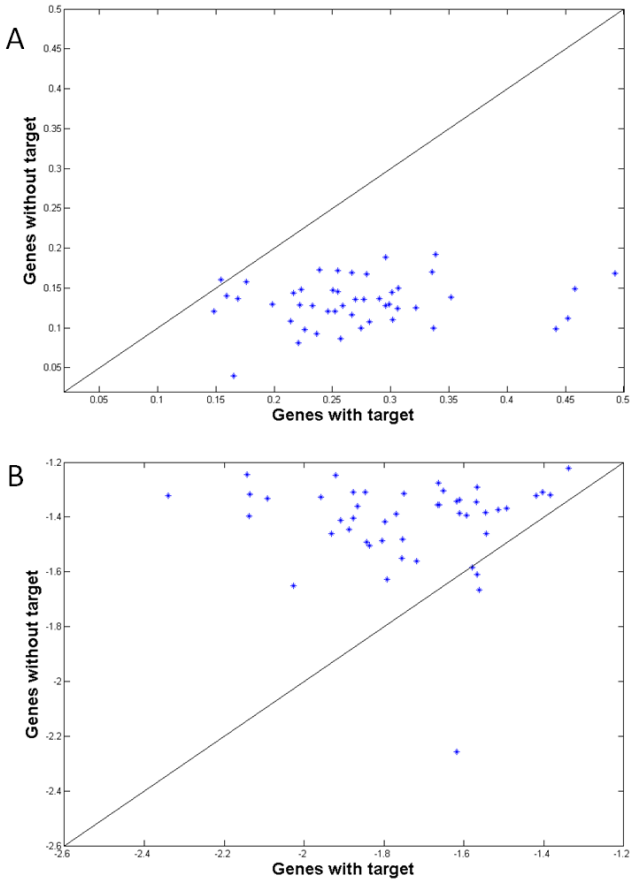

Figure 2: The importance of miR targets in the 3'UTR. In each microarray experiment two groups of genes were compared: genes with and without targets for the overexpressed miR. A: percentage of downregulated genes in the two groups. B: the average fold change (log2 scale) of genes in the two groups.

Conserved miR targets are more functional than non conserved miR targets

The mere existence of a miR binding site sequence inside the 3'UTR of a gene does not necessarily imply that the gene will constitute a functional target of the miR. A common hallmark of target's authentic functionality is its conservation[73]. Therefore, I next compared between genes that contain a conserved vs. non conserved binding motif for each tested miR. Figure 3 shows that the group of genes with the conserved motif exhibited a significantly larger percentage of genes that were downregulated when the miR was overexpressed (Student's t-test P-value of 6.5e-14), as well as a higher average fold reduction in expression (Student's t-test P-value of 8.2e-5) compared to genes that contain the same motif yet that motif is not conserved in other mammals.

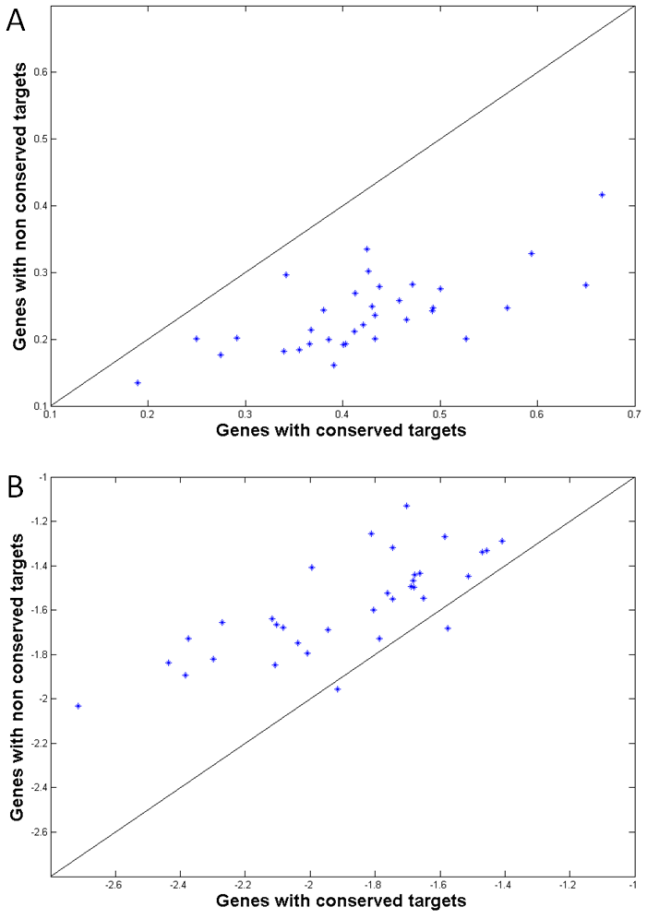

Figure 3: The importance of conserved miR targets in the 3'UTR. In each microarray experiment two groups of genes were compared: those that contain a conserved target for the overexpressed miR and those with a non conserved miR target. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

Multiple appearances of the miR target contribute to the target's functionality

The number of times a target appears in the 3'UTR of a particular mRNA may be a crucial feature of the target's functionality. The contribution of this feature can be at two levels: cooperativity in the binding of the miRNAs and cooperativity in the function of the miRNAs after binding. Brenneck et al. showed that when the *hid* gene in the drosophila had three instead of five binding sites for the miR encoded by the *bantam* gene, its regulation was weaker[74]. Doench et al. compared between one binding site per gene and more than one when using siRNAs, and found that it was a non-linear connection, suggesting that the binding sites of miRNAs as well may function cooperatively[75].

As shown in Figure 4, I also found that genes with more than one binding site motif for a given miR are more downregulated than genes with only one motif for that miR, in downregulation percentage as well as average fold change of the downregulated genes (Student's t-test P-value of 6.3e-14 and 6.6e-6 respectively).

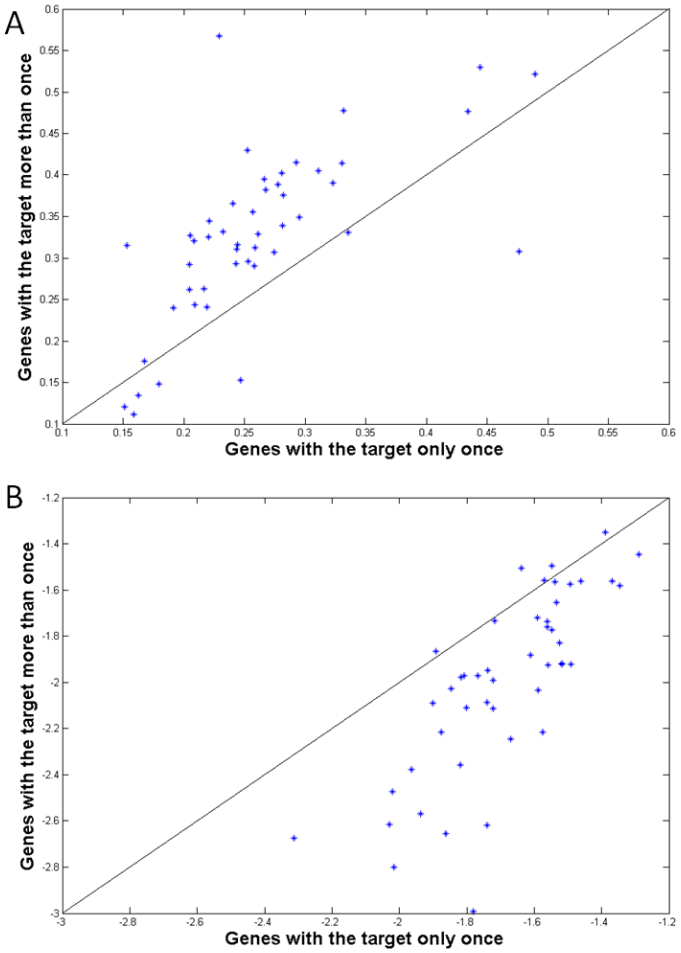

Figure 4: The importance of the number of appearances of the miR target in the 3'UTR. In each microarray experiment two groups of genes were compared: those containing the overexpressed miR target only once and those harboring the target multiple times. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

The potential genomic interplay between Alus and miRs

As Alus are so abundant in the human genome, the potential effect of functional miR targets inside them might be substantial and impact crucial cellular processes. To evaluate the potential effect of Alu insertions in the human genome on miR targeting I first examined how many predicted miR targets (perfect 7-mers) appear inside Alu sequences that reside in the 3'UTRs of human genes. This analysis revealed that 16% of the human genes contain at least one Alu in their 3'UTR. A total of 4927 Alu sequences that reside within the 3'UTRs present 94,785 potential miR targets, and in particular 28,829 targets of the 401 miRs that are conserved among mammals. Notably, there is a total of 3,088 predicted targets for the most conserved miRs in the animal kingdom (74 miRs)[76]. Hence, the potential impact of miR targeting Alu sequences may be substantial.

miR binding sites inside Alus are less functional

To discriminate between the effect of miR sites residing inside versus outside Alus I defined two groups of genes for each miR in the dataset: those that harbor the binding motifs only inside Alus and those that harbor the target only outside Alus (Table 1). As shown in Figure 5, I found that the percentage of downregulated genes is significantly greater when the targets reside outside Alus as compared to when they are located only inside Alus (Student's t-test P-value of 1.1e-7). The same trend appears in our alternative measure of miR overexpression efficacy, namely the average extent of downregulation per gene: the fold reduction is significantly higher for the genes that contain the miR site outside Alus (Student's t-test P-value of 0.004).

Since Alus are primate-specific and therefore any motif inside an Alu is by default non-conserved in mammals, I wanted to rule out the possibility that the lower functionality of the miR binding sites inside Alus is not unique to Alus but rather is a mere reflection of the fact that they are not conserved. I therefore compared the effect of Alu-hosted miR targets to non-conserved targets residing outside Alus. As shown in Figure 6, this analysis revealed that miR binding sites inside Alus are still less functional even when compared to non-conserved miR binding sites outside Alus (Student's t-test P-values of 4.6e-4 and 2.4e-6 respectively).

I next wished to assess the impact of miR targets residing only within Alus, relative to transcripts that don't harbor any target at all for that miR. The results (Figure 7) indicated that, on average, the two groups of transcripts exhibited similar responses to the overexpression of the relevant miR

(Student's t-test P-values of 0.67 and 0.93 respectively), suggesting that the vast majority of miR sites within Alus are probably non functional.

In the above analysis, I included only targets that represent a perfect 7-mer match to the miR seed, and therefore are expected in principle to be effective. This implies that there must exist a molecular mechanism by which these targets are "hidden" from the miR recognition machinery. I hypothesized that the apparent lack of function of Alu-contained miR sites may stem from specific genomic features associated with such sites. I thus next set out to identify such features that may potentially enable miR binding sites within Alus to avoid recognition by the RISC machinery.

miR	target	num genes with target only in alu	num genes with target only out alu	num genes with both in and out alu
hsa-let-7b	CTACCTC	67	1245	14
hsa-miR-1	ACATTCC	5	1630	1
hsa-miR-103	ATGCTGC	12	1741	1
hsa-miR-106b	GCACTTT	826	1894	202
hsa-miR-107	ATGCTGC	12	1741	1
hsa-miR-122a	ACACTCC	51	1068	8
hsa-miR-124	GTGCCTT	9	2024	5
hsa-miR-132	GACTGTT	6	1485	2
hsa-miR-133a	GGACCAA	1	1077	0
hsa-miR-142	ACACTAC	18	723	1
hsa-miR-148b	TGCACTG	62	1900	14
hsa-miR-155	AGCATTA	1	1432	1
hsa-miR-15a	TGCTGCT	9	3086	2
hsa-miR-15b	TGCTGCT	9	3086	2
hsa-miR-16	TGCTGCT	9	3086	2
hsa-miR-181	TGAATGT	3	2690	0
hsa-miR-195	TGCTGCT	9	3086	2
hsa-miR-30a	TGTTTAC	2	2235	1
hsa-miR-34a	CACTGCC	41	2244	8
hsa-miR-34c	TAGTGAT	14	1160	2
hsa-miR-373	AGCACTT	830	1827	215
hsa-miR-7	GTCTTCC	15	1584	4
hsa-miR-9	ACCAAAG	10	1771	2

Table 1: Numbers of genes containing binding sites for each of the miRs included in the Khan dataset. For each miR the table shows the number of genes in the genome that contain the miR target in its 3'UTR inside and outside Alus.

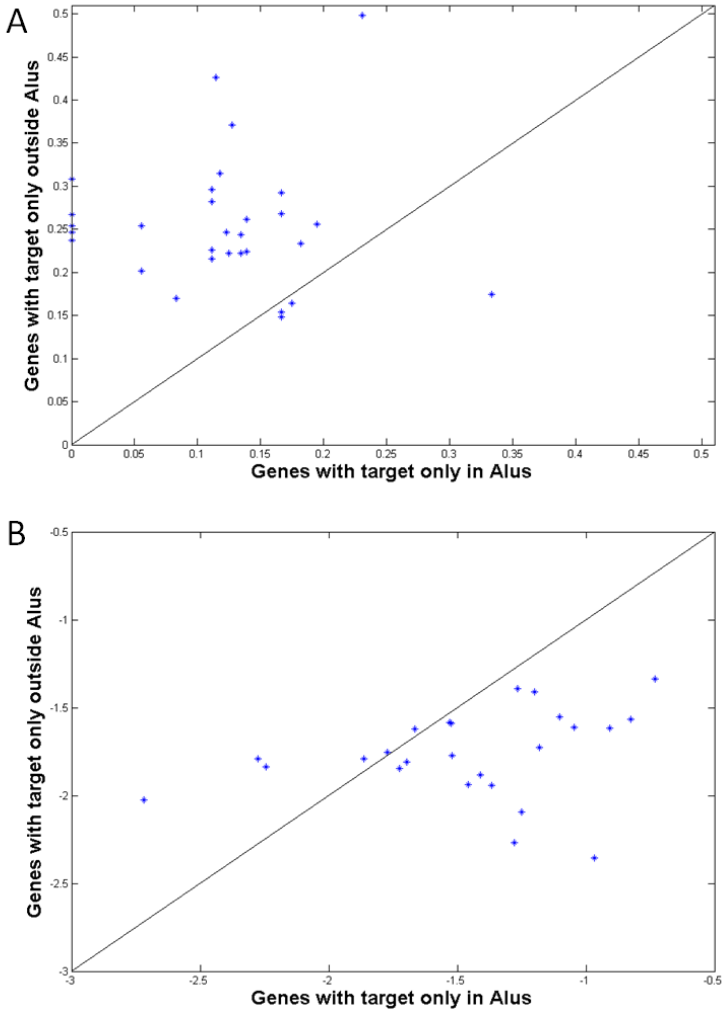

Figure 5: miR targets inside Alus are less functional than outside Alus. For each microarray experiment two groups of genes were compared: those with the overexpressed miR target only inside Alus and those with the target only outside Alus. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

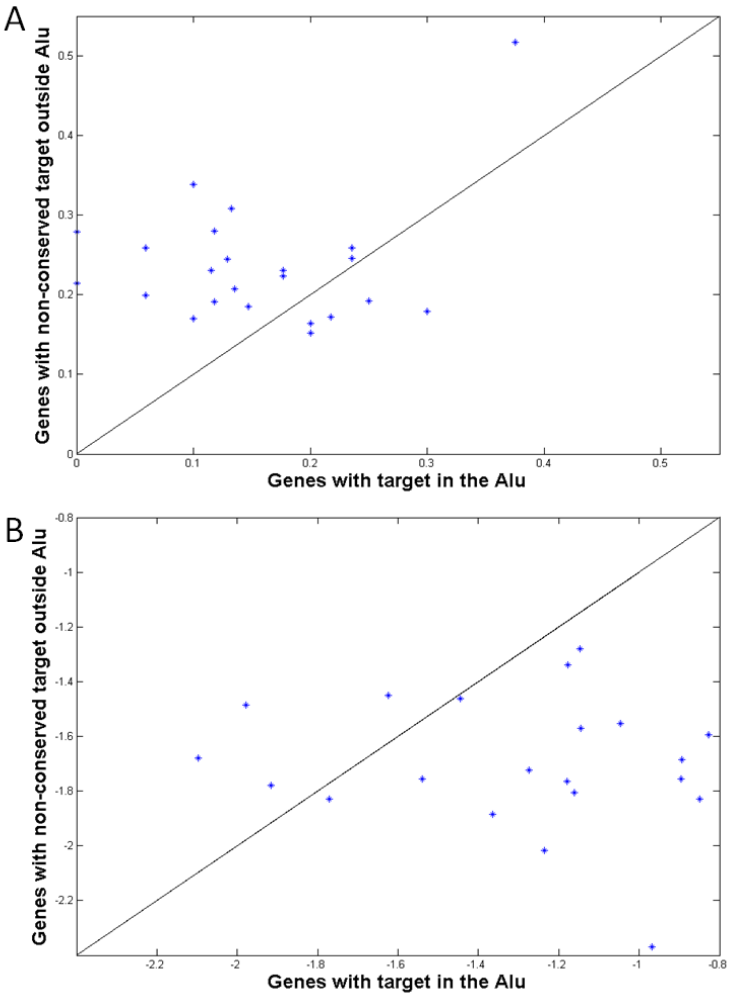

Figure 6: miR targets inside Alus are less functional than non-conserved targets outside Alus. For each microarray experiment two groups of genes were compared: those with the overexpressed miR target only inside Alus and those with a non-conserved target outside Alus. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

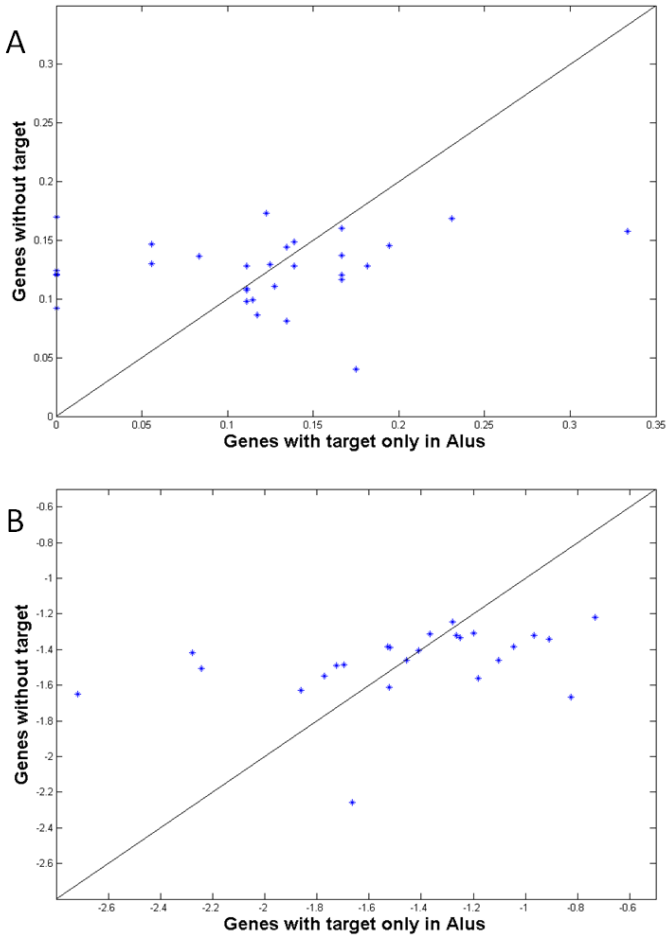

Figure 7: miR targets inside Alus tend to be non-functional. For each microarray experiment genes with a target for the overexpressed miR only inside Alus were compared with those that don't harbor any target at all for this miR within their 3'UTR. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

Mutually distinct 3' UTR territories of conserved miR sites and Alu repeats

The first parameter that I explored in order to elucidate the lack of regulatory function of miR binding sites within Alus is their location within the 3'UTR. It was previously shown that conserved and functional miR targets tend to reside preferentially near both ends of the 3'UTR rather than in the middle part of the 3'UTR[5]. The analysis that I performed on the Khan dataset (Figure 8) recapitulated the above conclusion, further showing that miR binding sites in the middle are less functional than those in the ends, (Student's t-test P-values of 0.03 and 0.17).

Another way of recapitulating this statement is by looking at the positions of conserved miR targets. Figure 9A shows that conserved miR binding sites are enriched near the two ends of the 3'UTR. Moreover, by analysis of 40,157 conserved targets, I found that these conserved sites are more abundant in the first and last 250 bp of the 3'UTR (Figure 10B, D). In contrast, non-conserved sites are equally represented throughout the 3'UTRs (Figure 9C).

I next examined whether the position of Alu sequences within the 3'UTR is random or is also restricted and biased. Interestingly, I found that Alus are depleted from the 3'UTR ends, showing an almost exact mirror image of the conserved miR sites' distribution. This phenomenon was apparent from analysis of the relative position (Figure 9B) as well as of the absolute position (Figure 10A, C), implying that Alus are selectively excluded from the first and last 250bp.

These results suggest that Alu insertions in 3' UTRs were tolerated during primate evolution only provided that they occurred away from the two 3' UTR ends. It is tempting to speculate that one reason for such Alu avoidance from the 3' UTR ends is that such territories are preserved for functional miR regulation that need not be distorted by Alu insertions.

In order to check whether this was the only explanation for the disfunctionality of miR binding sites inside Alus, I inspected the Khan dataset for genes with the target only inside Alus vs. genes with the target outside Alus but nevertheless located in the middle of the 3'UTR rather than near its ends. As can be seen in Figure 11, targets in the middle of the 3'UTR were still found to be more functional than those inside Alus (Student's t-test P-values of 7.9e-11 and 0.0026). Therefore, there must exist at least one additional mechanism besides the location of the miR binding sites which prevents targets inside Alus from being functional. This conjecture is further supported by the fact that, although most Alus are indeed located towards the middle of the 3'UTR, hundreds of Alu sites are nevertheless present in 3' UTR ends of genes,.

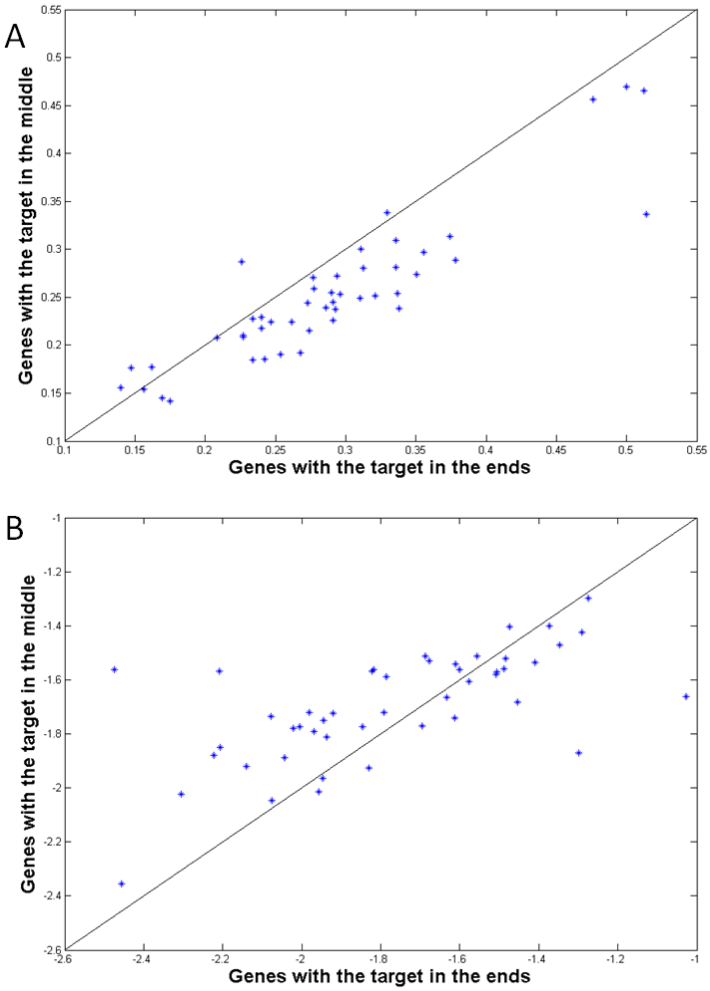

Figure 8: miR targets in the middle of the 3'UTR are less functional than near the 3'UTR ends. For each microarray experiment genes with the overexpressed miR target only near the ends (first and last 250 bases of the 3'UTR) were compared to genes with the target only in the middle of the 3'UTR. Only genes with 3'UTRs longer than 1000bp were included in the analysis. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

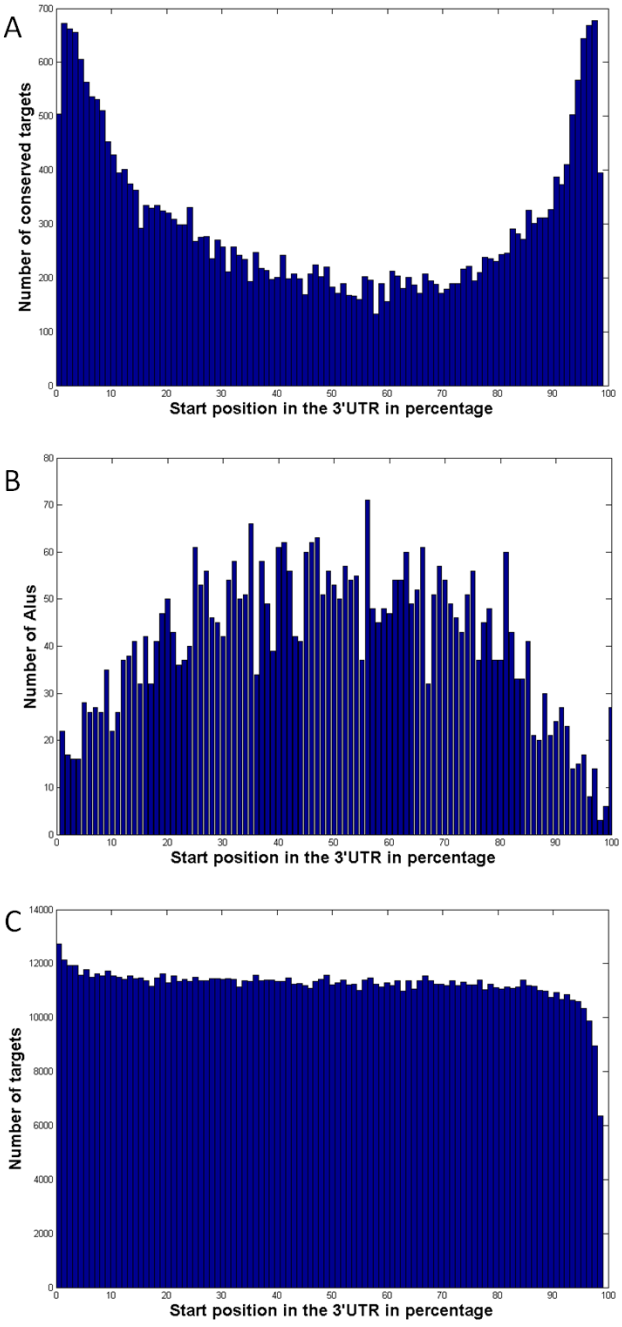

Figure 9: Conserved miR targets tend to reside near the 3'UTR ends, while Alu elements reside preferentially in the middle of the 3'UTR. The location (shown as cumulative percentage) of all miR targets and Alus in genes with 3'UTR above 1000 bases were calculated. A: conserved miR targets. B: Alu sequences. C: conserved and non-conserved miR targets together.

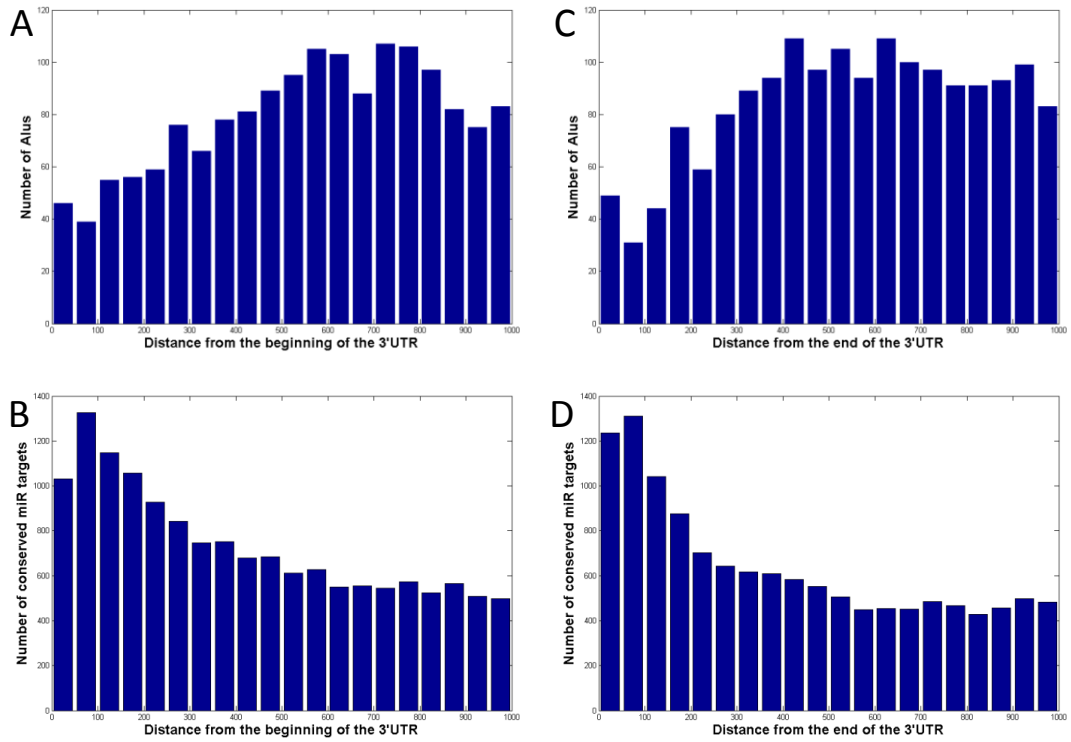

Figure 10: Conserved miR targets tend to reside near the 3'UTR ends, while Alu elements reside preferentially in the middle of the 3'UTR. The absolute distance from the beginning and end of the 3'UTR was calculated for all miR targets and Alus in genes with 3'UTRs longer than 1000 bases. A, C: Alu sequences. B, D: conserved miR targets.

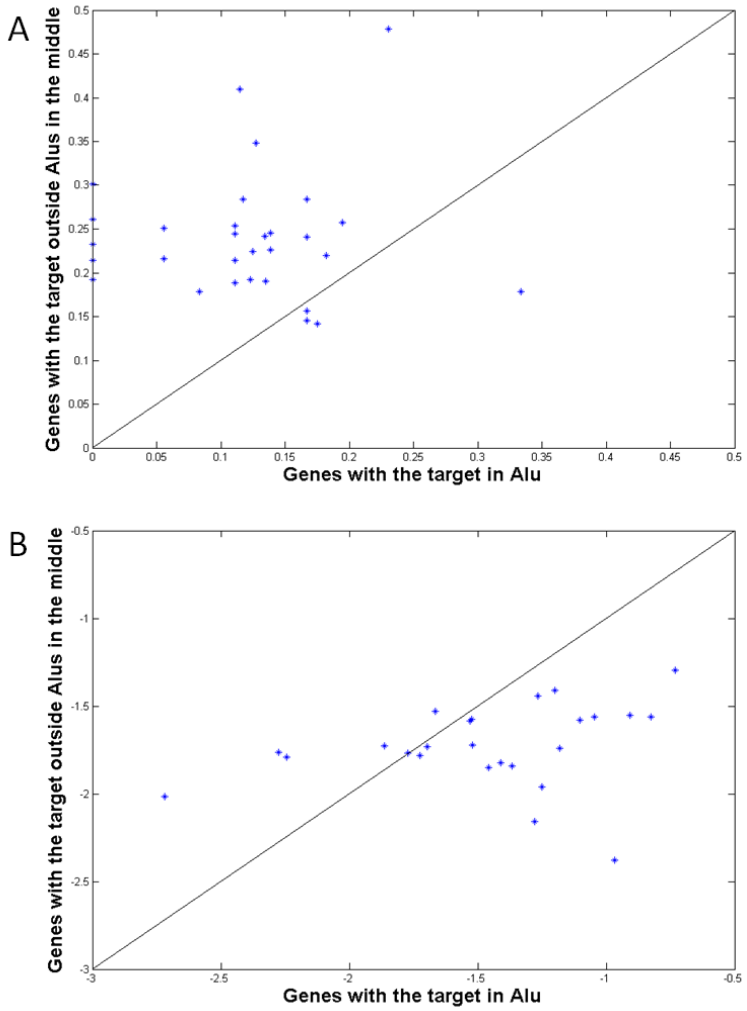

Figure 11: miR targets in the middle of the 3'UTR residing outside Alus are more functional than targets inside Alus. In each microarray experiment genes with the overexpressed miR target only in Alus were compared with genes harboring the target only in the middle of the 3'UTR (middle is all the 3'UTR but the first and last 250 bases). Only genes with 3'UTRs above 1000bp were taken into the analysis. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

miR targets inside Alus have more extensive secondary structure than targets outside Alus

It was previously shown that the structure and folding energy of miR targets and their surroundings are very important for their functionality[77]. Specifically, Segal and colleagues showed that targets within tight secondary structures are less functional[77].

I therefore simulated the structure of the full genome 3'UTRs, and looked at the mean energy profile of genes with vs. without conserved targets in specific areas. As can be seen in figure 12, genes that have conserved binding sites at a given location along the 3' UTR tend to have a significantly less tight secondary structure in that region compared to genes without any conserved miR site at that region, and this conclusion holds for all possible locations along the 3' UTR.

Another confirmation that the structure is important for the functionality of the miR binding site can be seen in figure 13, where I looked at the percentage of downregulated genes in the Khan dataset, and separated genes with structured miR binding sites from genes with un-structured miR binding sites. It can be seen that there is a significant difference between the two groups, and un-structured miR binding sites are more functional than structured miR binding sites (Student's t-test P-values of 0.06 and 0.03).

As tight secondary structure around the target site might serve as an additional avoidance mechanism, I examined the folding energy around targets inside Alus. As illustrated in figure 14, targets inside Alus tend to reside within more tight structures, as compared with conserved targets, an observation that is compatible with the known high secondary structure content of RNA encoded by Alu repeats[78-79]. Yet, this could not be the main mechanism for avoidance, as there are numerous conserved targets whose extent of predicted folding is comparable to that of binding sites inside Alus (see histogram), and these conserved targets are nevertheless functional, as exemplified by our analysis of Khan's data.

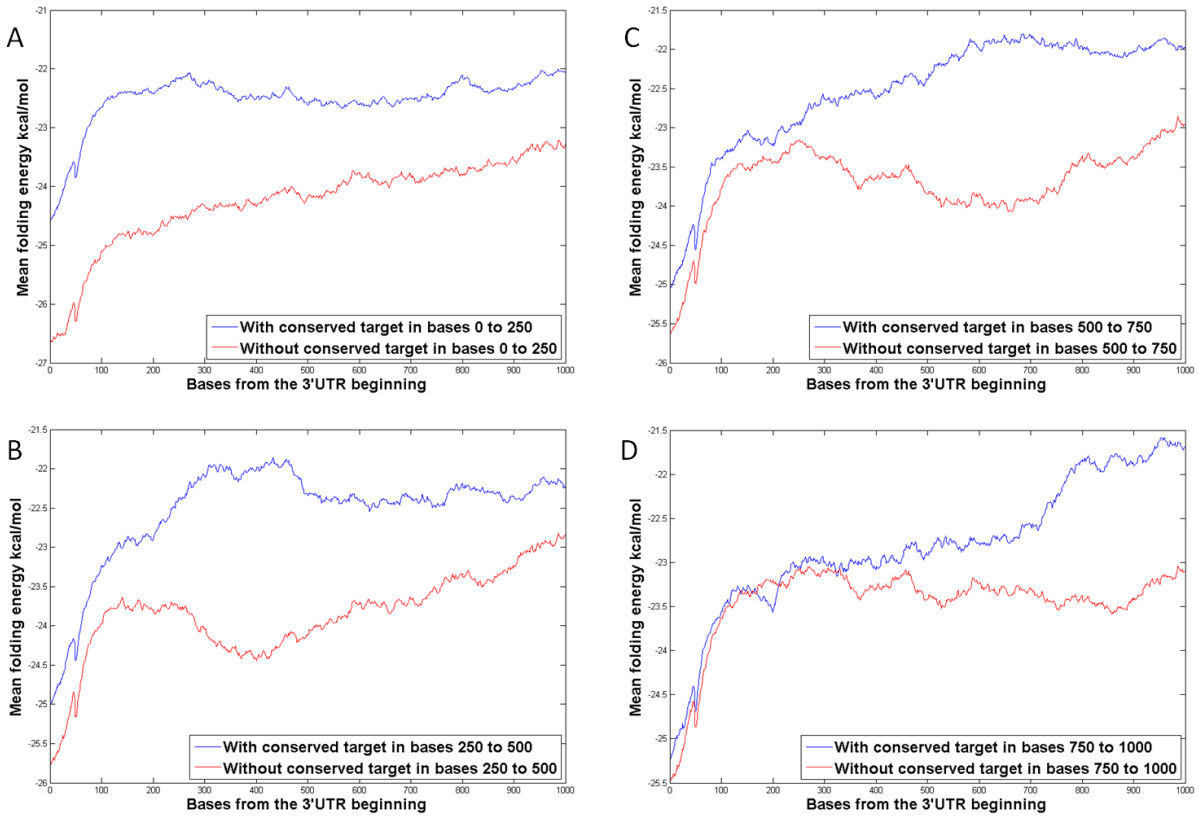

Figure 12: Mean folding energy around conserved targets is higher than around non-conserved targets. The mean folding energy of all genes with 3'UTR>1000 bp was simulated in Matlab. In figures A-D the mean folding energy in the first 1000 bases of the 3'UTR is presented for two groups of genes: those with conserved targets in a specific area, and those without conserved targets in a specific area.

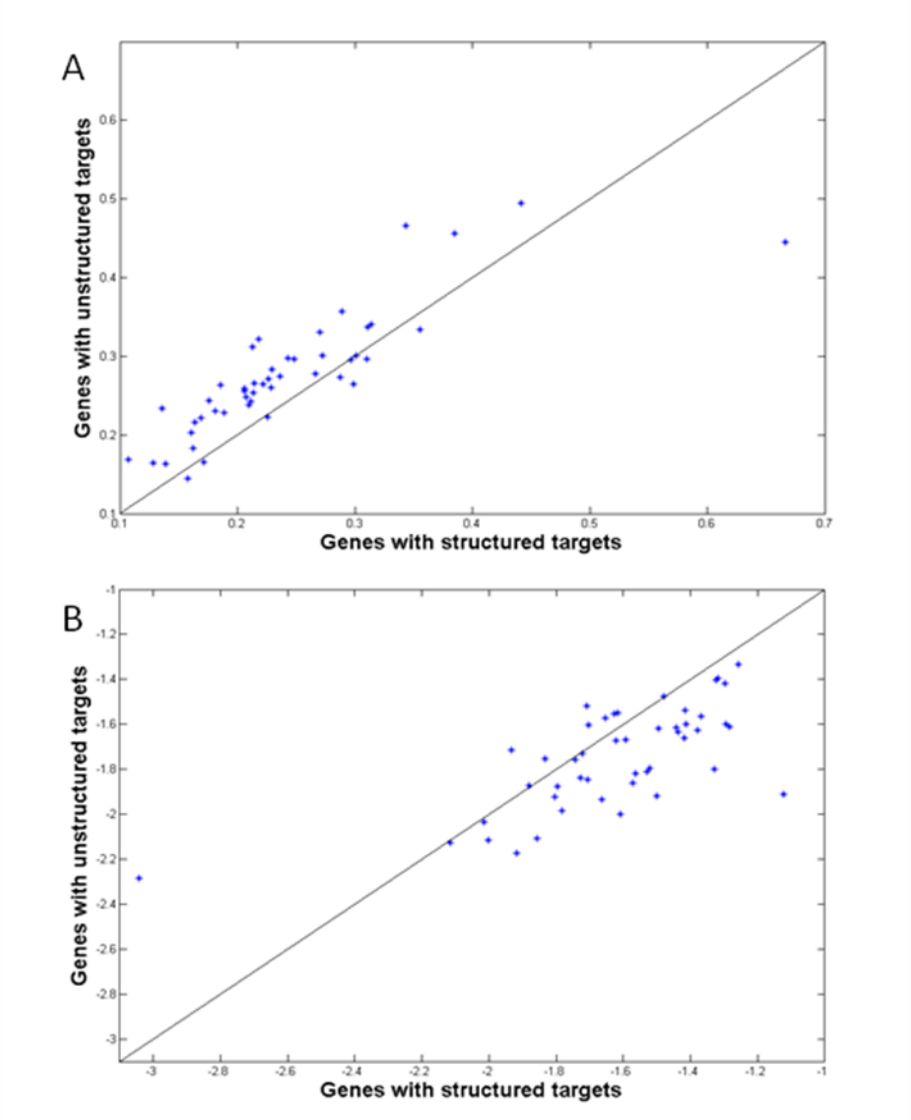

Figure 13: Importance of RNA structure for miR binding site functionality. The structure of all targets was simulated in Matlab. A target was defined as structured when more than 80% of its sequence was predicted to undergo base pairing. A: percentage of downregulated genes in the two groups. B: average fold change (log2 scale) of genes in the two groups.

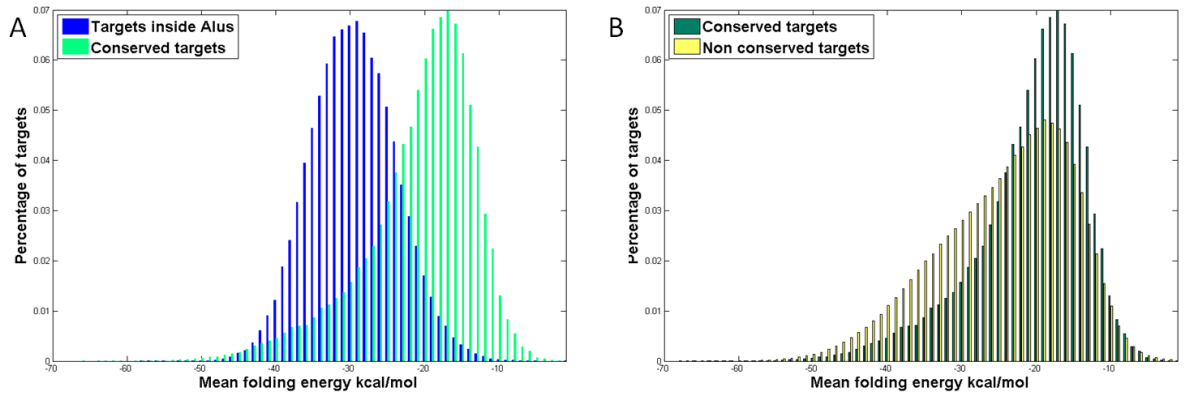

Figure 14: Targets inside Alus have a tighter structure than conserved targets. The mean folding energy of all the miR targets was calculated using Matlab. A: mean folding energy of targets inside Alus vs. conserved targets. B: mean folding energy of conserved vs. non-conserved targets. B: mean folding energy of conserved vs. non-conserved targets.

miR targets inside Alus appear to be subject to RNA editing

It was previously shown that Alu sequences in the human transcriptome are subject to extensive RNA editing[80-84], modifying adenosines to inosines. Since inosines are recognized as guanosines, such alterations can diminish the complementarity between a miR's seed and its binding site within Alus.

I first examined the miR targets that do not contain an A in their RNA sequences, as they cannot be subject to editing; out of 18 miR binding sites in the collection, 3 (16%) have no A in the respective seven-mer binding site.

When examining only genes with putative miRNA binding sites outside Alus, there was no difference in the response to miRNA overexpression between targets that contain an A in the motif and those that do not (Figure 15A, student t-test 0.3). In contrast, among genes with targets only inside Alus, potential miRNA targets containing at least one A were found to be less functional than those without any A (Figure 15B, student t-test 0.02).

These results suggest that RNA editing is an additional mechanism through which miR targets that contain adenosines and reside within Alus avoid miR recognition.

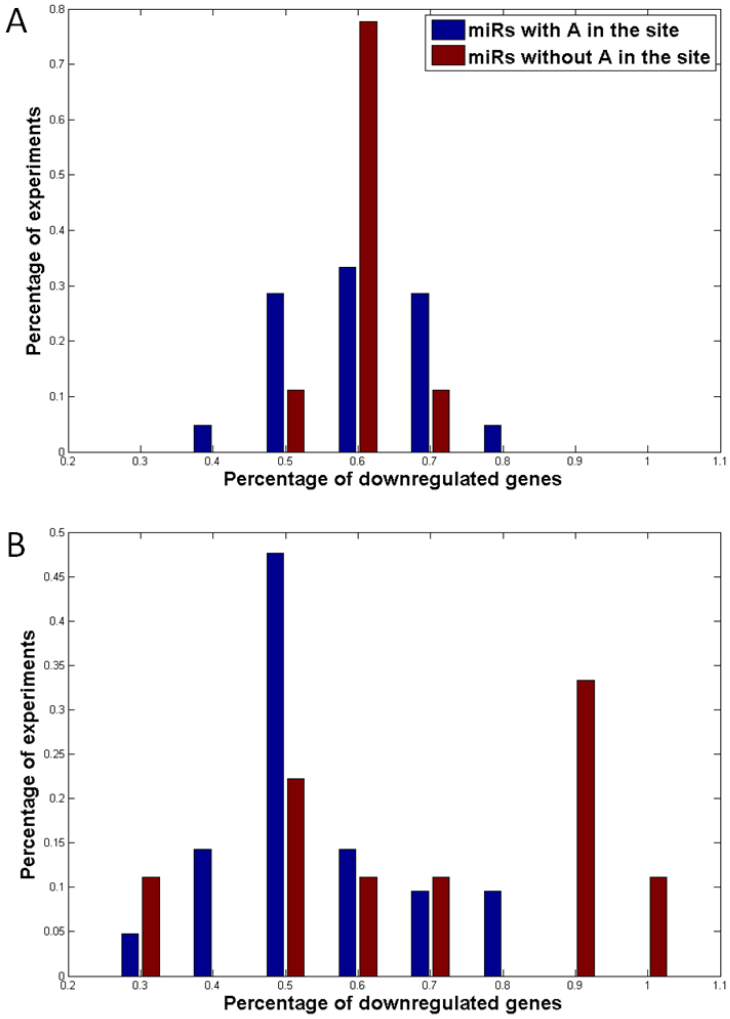

Figure 15: miR targets inside Alus may be masked by RNA editing. The group of microarray experiments were divided to those where the miR target contains an A and those where the miR target does not contain an A. A: percentage of downregulatted genes among those that contain the miR target only outside Alus. B: percentage of downregulated genes among those that contain the miR target only within Alus.

Analysis of the avoidance means of all targets of conserved miRs

After finding three potential mechanisms by which miR binding sites inside Alus can possibly avoid the miR machinery, I next wished to evaluate the full effect of those mechanisms and the interactions between them. As can be seen in figure 16, most targets within Alus possess at least one potential mechanism of avoidance. Most targets have at least two, and a large proportion have all three mechanisms.

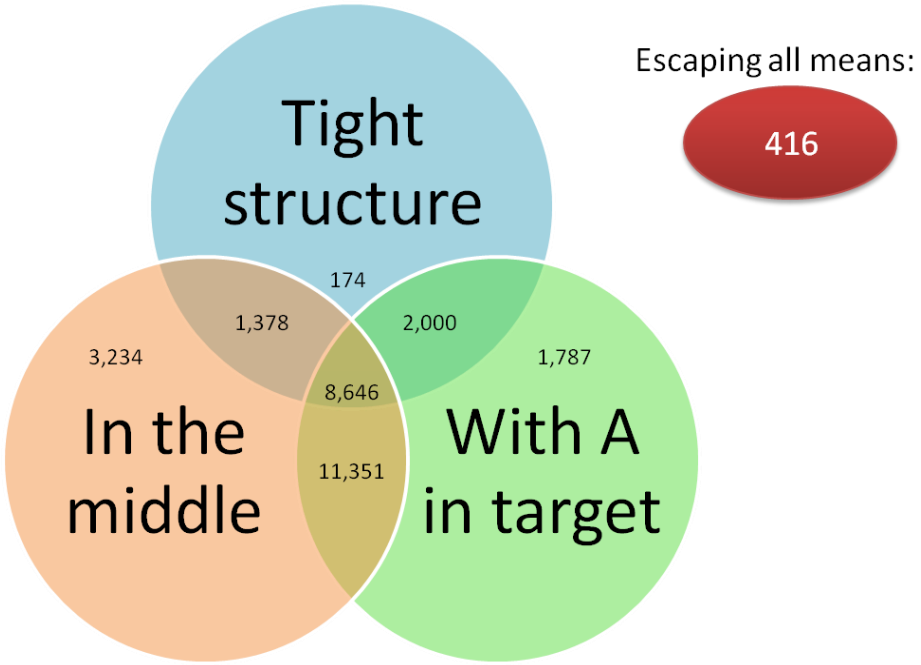

Figure 16: Analysis of the avoidance mechanisms of all targets of conserved miRs. All targets of conserved miRs (in mammals) within 3'UTRs that are at least 1000bp long were analyzed for location, structure and nucleotide composition.

Avoidance mechanisms of B1 repeats in the mouse genome

Transposable elements are active in most animal genomes; therefore introduction of novel miR targets via transposition can occur also in other species. To support our results in the human genome, I investigated this issue in the mouse genome.

Although Alu repeats are primate-specific, there are repeats similar to Alus in the mouse genome, namely, in the SINE family – B1 repeats. As is the case for Alus, B1 repeats emerged from the ancestral 7SL RNA gene[85]. The B1 repeats are less widespread (comprising only 2.7% of the mouse genome [8,86]) and are shorter (~ 140 bp)[87].

In the mouse genome, 8.3% of the genes contain at least one B1 repeat in their 3'UTR. The 1962 B1 sequences that reside within the 3'UTRs present 14,372 potential miR targets (perfect 7-mers). Consequently, the potential effect of miR targeting B1 is less substantial than in Alus in the human genome, but may still be meaningful.

RNA derived from B1 repeats is less likely to be subject to RNA editing, since their number is significantly smaller, therefore it is less likely of them to be close and in opposite direction to one another and form a secondary structure. I therefore focused on analyses of location and structure

With regard to the location within the 3'UTR, mouse B1 repeats show similar trends as the human Alus. B1 repeats tend to avoid the ends of the 3'UTR, predominantly the beginning, showing a trend opposite to the conserved mouse miR targets (Figure 17).

As expected from their common origin, miR targets within B1 repeats (like the ones within Alus) show lower folding energies in their immediate region, while conserved mouse miR targets show higher folding energies (Figure 18).

In conclusion, both avoidance mechanisms -mutual exclusion in the location and less accessible mRNA secondary structure - are apparent in the mouse genome, supporting their importance in avoiding the introduction of new miR targets.

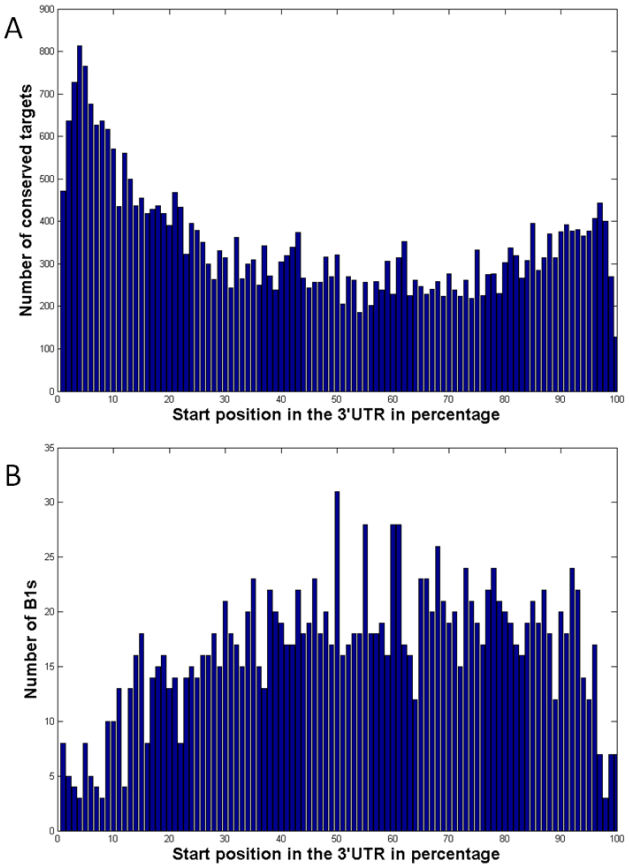

Figure 17: Repeats and miR targets reside in mutually distinct 3' UTR territories in the mouse genome. A: Relative positions along the 3'UTR of all conserved miR targets in 3'UTRs longer than 1000 bases. B: Relative start position along the 3'UTR of all B1 repeats in 3'UTRs longer than 1000 bases.

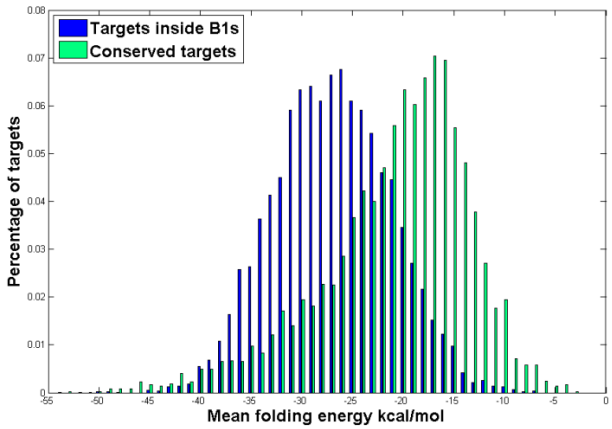

Figure 18: Targets inside B1 repeats have a tighter mRNA secondary structure than conserved targets. The mean folding energy of all the miR targets inside B1 sequences and conserved targets in the mouse genome was calculated using Matlab.

miR-661 and miR-339-5p in the p53 network

In the first part of this work I concentrated on the global effect of miR targets inside Alus. I showed that in general, miR targets inside Alus are being masked by different mechanisms in the cell, and that they are less functional than miR targets outside Alus.

However, it is entirely possible that a small fraction of those evolutionarily novel miR sites inside Alus are functional, and introduce new regulations to the cell.

In this part of my work my aim is to concentrate at the highly important p53 pathway, to find miRs that downregulate multiple components in the pathway, and specifically look at miR targets inside Alus in those components.

microRNA target prediction in the p53 network, guided by multiple appearances per target

As described in the Khan dataset analysis, the number of times a miR site appears in the 3'UTR of a particular mRNA may be a crucial feature of the site's functionality. The contribution of this feature can be at two levels: cooperativity in the binding of the miRNAs and cooperativity in the function of the miRNAs after binding. On the statistical basis, multiple occurrences of the same site in a given UTR might reduce the chance of false positive prediction. Based on these hypotheses, I decided to perform a search for miRNA targets in the p53 network, focusing on the multiple appearances feature. To that end, I scanned all possible nucleotide k-mers (for <=7k<=11) for the presence of each of them within the 3'UTRs of human genes' associated with the p53 network, as described in KEGG (http://www.genome.jp/kegg/). I devised two computational tests to find k-mers that tend to (1) appear multiple times in 3'UTRs of the same genes, or (2) multiple times in the 3' UTRs of multiple genes belonging to the p53 network. In those tests I considered k-mers that can serve as potential targets for known miRNAs, as well as on those that are not associated with a known miRNA. The first test was a binomial test, which gave a p-value for each pair that consists of a gene and a kmer for the hypothesis that the number of appearances of the k-mer in the gene's UTR reflects the genome average for that k-mer. In this test n was the number of times the kmer could appear in a gene's 3'UTR (assessed as the length of the 3'UTR divided by the length of the kmer), and p was the average score of appearances of the kmer in the 3'UTRs of all genes. False Discovery Rate (FDR) was applied and set on a q-value of 10%. The second test was the "gap test": the gap score is the number of total appearances of the kmer in the 3'UTRs of the p53 network genes minus the number of genes the kmer appears in the network. The score of each kmer was its gap score minus its average gap score in the

controls. I created three controls: 1. a group of 700 random genes that have the same 3'UTR length distribution as the p53 network genes 2. a shuffle of the genes' 3'UTRS and 3. a shuffle of the kmer. I considered the top 1000 kmers that had the largest difference between the gap score in the p53 network genes and in the controls.

Among the kmers that passed both tests (the binomial and the gap test) was the palindrome CCAGGCTGG, which appears in ~2550 genes in the genome. Among all gens in the genome, it received binomial scores that passed FDR 10% for 37 genes, among them are MDM2 and MDM4. In the p53 network its average appearances (in genes it appears in) is 2.7, while in the full genome its average is 1.79. This palindrome did not match any known human miRNA, but I noted that addition of one C at its beginning converted it into a miR-661 target (CCCAGGC). Remarkably, a perfect putative target of miR-661 appears 5 times in *MDM2* mRNA and 9 times in *MDM4* mRNA. This was a very intriguing observation, since both MDM2 and MDM4 are major regulators of p53. Specifically these two factors work together to promote p53 degradation and inactivation. Of particular relevance to the concepts discussed earlier in this thesis, all the miR-661 targets in *MDM2* mRNA and all but one in *MDM4* reside within Alus. Had we masked 3' UTRs from repetitive elements prior to the computational screen we would have missed this site.

In addition to miR-661, the target of miR-339-5p also appeared to have good scores. It was in the top 1000 gap score kmers and had low p-values in the binomial test, although it didn't pass the FDR test. The miR-339-5p target in *MDM4* and MDM2 scored p-values of 0.0076 and $1.74E^{-4}$, respectively. In *MDM4* mRNA, miR-339-5p has 4 targets and 3 in *MDM2*.

I decided to focus on those two miRs to see their effect on MDM2 and MDM4 expression in living cells.

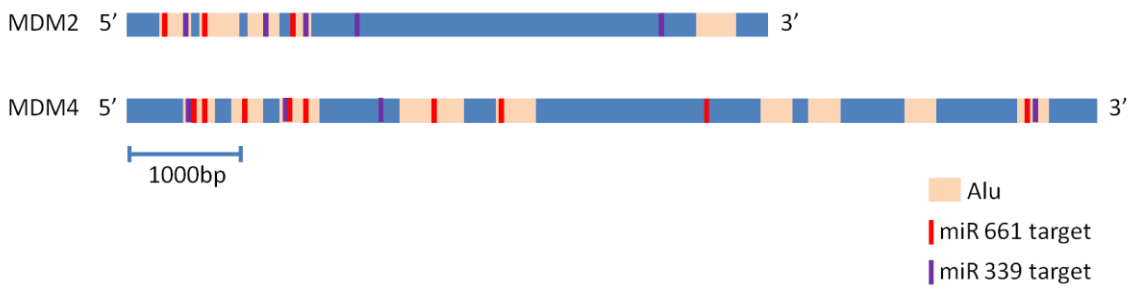

Figure 19: miR-661 and miR-339-5p have multiple targets in *MDM2* and *MDM4* mRNA. The 3'UTRs of MDM2 and MDM4 are presented, with Alus and targets for miR-661 and miR-339-5p. Targets are perfect matching 7-mers.

miR-661 and miR-339-5p downregulate MDM2 and MDM4

In order to investigate whether miR-661 and miR-339-5p can indeed target *MDM2* and *MDM4*, I first checked their impact on the respective mRNA levels. MCF7 cells were transfected with the specific miRs vs. a miR control (which is not supposed to target these mRNAs) for 48 hours. As can be seen in figure 20A, MDM2 mRNA levels were suppressed by both miRs: miR-661 and miR-339-5p caused a reduction of ~20% and ~25%, respectively, in *MDM2* mRNA. The levels of *MDM4* mRNA were not significantly affected by these miRs, and no significant change was also seen in the levels of *p53* mRNA. However, the mRNA level of the p53 transcriptional target *p21* was strongly upregulated upon both miR-661 and miR-339-5p overexpression, suggesting that p53 has undergone activation.

Next, I looked at the protein levels of MDM2 and MDM4 after overexpression of miR-661 and miR-339-5p. As can be seen in figure 20B, the levels of MDM2 were downregulated both by miR-661 and miR-339-5p, while MDM4 was downregulated only by miR-661. In contrast, p53 protein levels were not affected.

At first glance, the lack of change in p53 protein levels was puzzling, since a decrease in MDM2 and MDM4 might be expected to slow-down p53 ubiquitination and therefore lead to p53 accumulation. However, in addition to promoting p53 protein degradation, MDM2 and MDM4 also inhibit the biochemical functions of p53, even in cases where there is no visible effect on p53 protein levels. It is thus conceivable that the increase in *p21* mRNA levels is indicative of an upregulation in the functionality of p53.

In conclusion, I show that miR-661 downregulates MDM2 at the mRNA and protein level and MDM4 only at the protein level. Furthermore, miR-339-5p targets MDM2, downregulating its mRNA and protein levels. This is consistent with the computational predictions discussed above.

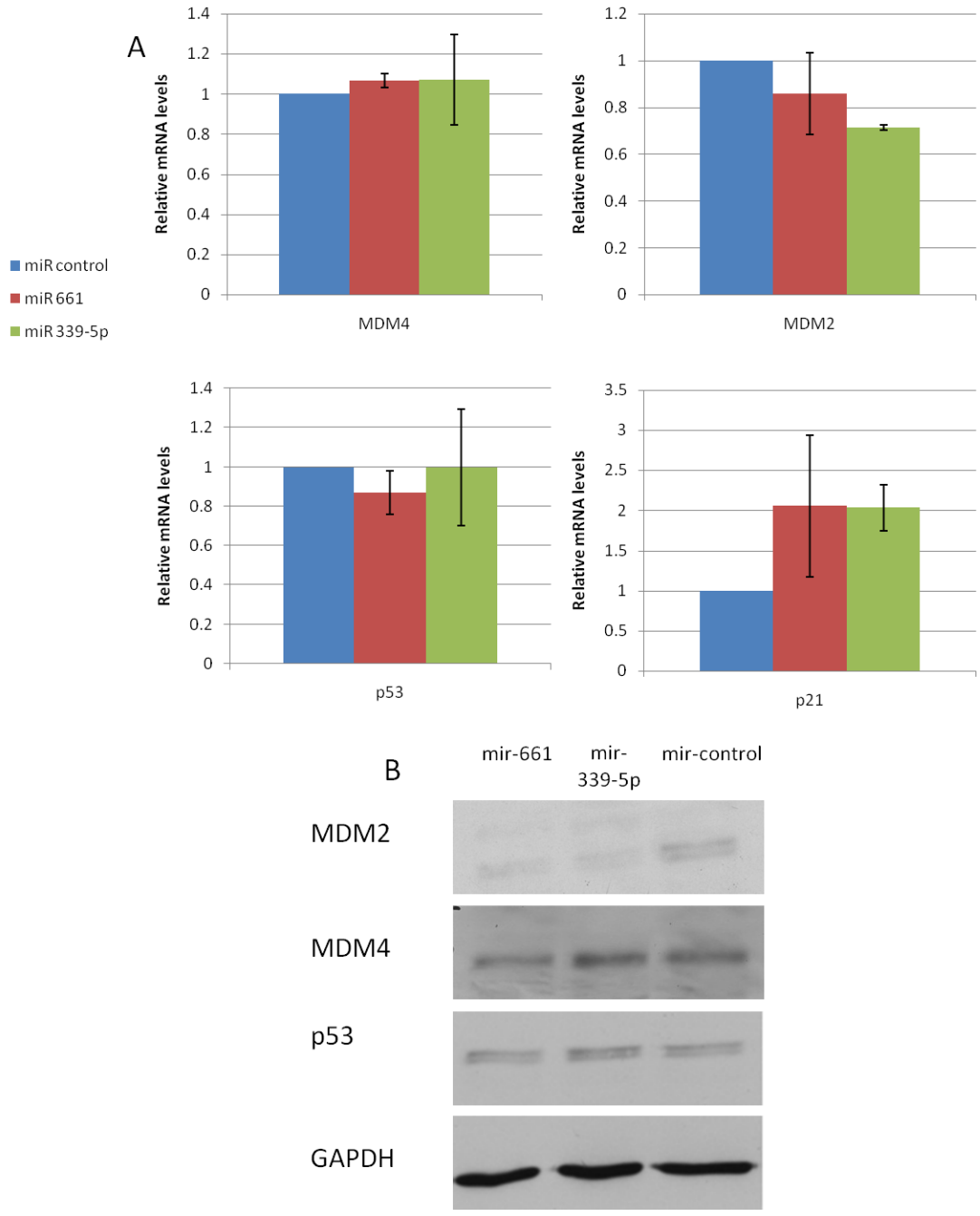

Figure 20: miR-661 and miR-339-5p downregulate MDM2 and MDM4. MCF7 cells were transfected with 20nM miR-661 mimic or miR-339-5p mimic or miR-control, and harvested for analysis 48 hours later. A: qRT-PCR analysis of *MDM4*, *MDM2*, *p53* and *p21* mRNA. Values were normalized to *GAPDH* mRNA in the same sample. B: Protein analysis was performed by Western blot with the indicated antibodies.

p53 reduces the effect of of miR 661 and miR 339 on MDM2

MDM2 levels are positively regulated by p53, thus forming a negative feedback loop. In order to assess more directly the effects of the two miRs on MDM2 levels, I therefore knocked down p53. MCF7 cells were transfected with miR-661, miR-339-5p or control miR, as well as with p53 siRNA (sip53) or control siRNA (siControl). Cells were harvested for analysis 48 h later. As shown in figure 21A, overexpression of miR-661 or miR-339-5p in p53-depleted cells led to a ~25% or ~61% decrease, respectively, in *MDM2* mRNA levels. This downregulation was substantially stronger than in cells transfected with control siRNA, confirming the expectation that the p53-MDM2 loop dampens part of the effects of those miRs on MDM2.

The regulation of MDM4 by each of the two miRs was not affected by p53 knockdown, as expected from the fact that, unlike MDM2, MDM4 is not regulated by p53.

Levels of *p21* mRNA were upregulated when the miRs were overexpressed, and this effect was reversed by p53 knockdown, confirming that the increase in *p21* was a result of upregulation of p53 activity.

Similar conclusions were reached by Western blot analysis of the relevant proteins (figure 21B). As expected and as seen at the mRNA levels, the downregulation of MDM2 by the miRs was more prominent after knocking down p53. In addition, there was an upregulation of p21 protein by the miRs, which was counteracted by p53 knockdown in line with the effects on the mRNA levels

In summary, the downregulation of MDM2 mRNA and protein by miR-661 and miR-339-5p is convincingly observed upon downregulation of p53, but is partly occluded when p53 is fully active, because the induced p53 causes a secondary induction of MDM2 via transcriptional activation of the *MDM2* gene.

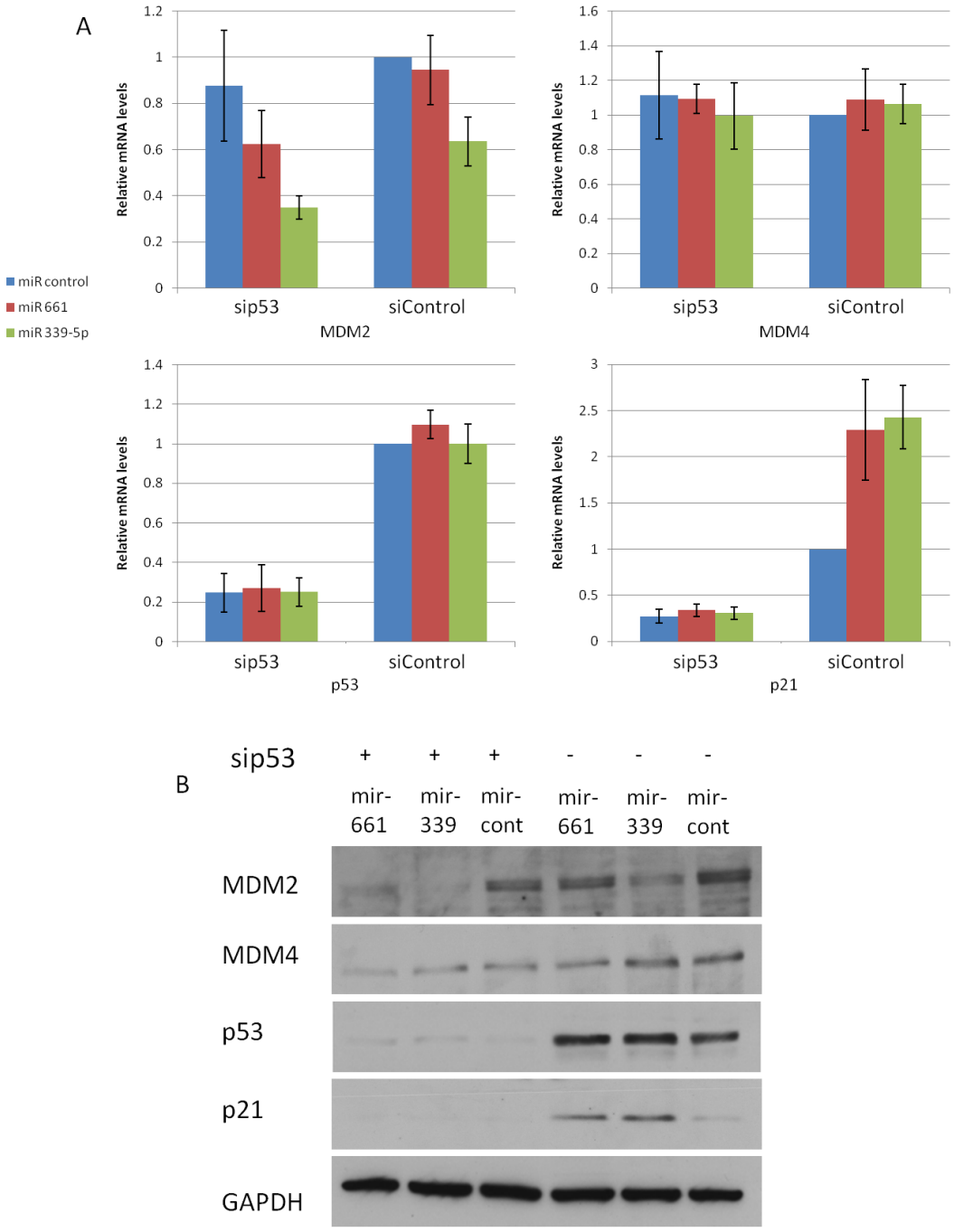

Figure 21: p53 depletion enhances the effect of miR-661 and miR-339-5p on MDM2 but not MDM4. MCF7 cells were transfected with 20nM miR-661 or miR-339-5p mimic or with miR-control, in combination with either 20nM p53 siRNA (sip53) or control siRNA (siControl) Cells were harvested for analysis 48 hours later. A: Total RNA was extracted and subjected to qRT-PCR analysis of *MDM4*, *MDM2*, *p53* and *p21* mRNA. Values were normalized to *GAPDH* mRNA in the same sample B: Cells were subjected to Western blot analysis with the indicated antibodies.

miR-661 overexpression induces activation of the *p21* gene promoter

I showed in previous sections that miR-661 overexpression causes an upregulation of p21 mRNA and protein, presumably via an increase in p53 activity.

In order to corroborate this conclusion, and to link it directly to enhanced transcriptional activity I looked at the effect of miR-661 on *p21* promoter activity. MCF7 cells were transfected with the miRs for 48h, with or without a plasmid containing the *p21* promoter, containing the p53 response elements, upstream of the luciferase gene. As a control, a version containing mutated p53 binding sites was transfected; this mutant is expected to not to be, regulated by p53. Luciferase levels were measured by luminescence.

As can be seen in figure 22, the activity of the wild type *p21* promoter was increased by ~2 fold as a result of miR-661 overexpression, while the activity of the mutant *p21* promoter was not affected. This result indicates that the increase of p21 after transfection of miR-661 is due to upregulation of p53 transcriptional activity.

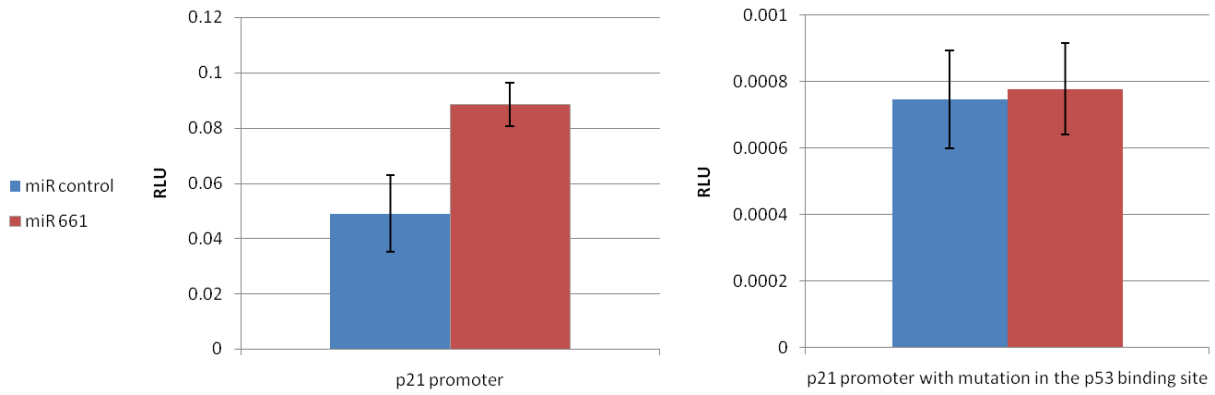

Figure 22: *p21* promoter activity is induced by miR-661 in a p53-dependent manner. MCF7 cells were transfected with 20 nM miR-661 or miR-control for 48 hours, followed by transfection of a luciferase reporter plasmid containing the wild type *p21* promoter or derivative thereof carrying mutations in the p53 binding sites. A plasmid expressing the Renilla luciferase gene was cotransfected and used as a normalization control.

miR-661 and miR-339-5p upregulate various p53 targets

p53 is a transcription factor with hundreds of targets. It may direct the cell to different outcomes such as senescence, growth arrest, or apoptosis. As established in previous sections, miR-661 and miR-339-5p are able to upregulate p53 activity by downregulating MDM2 and MDM4 levels. To further investigate the outcome of this upregulation, I monitored the endogenous mRNA levels of different p53 transcriptional targets.

As seen in figure 23, the levels of several such p53 targets were increased as a result of miR-661 and miR-339-5p overexpression, and the proteins encoded by these p53 target genes are involved in mediating diverse biochemical and biological effects of p53: CD95 encodes the Fas receptor that plays an important role in apoptosis, Btg2 controls the cell cycle, and Wig1 is a growth inhibitory zinc finger protein. Importantly, the effect of miR-661 and miR-339-5p on all those genes was abrogated upon knocking down p53.

The transcriptional upregulation of all these p53 targets by miR-661 and miR-339 implicates both miRs as potential tumor suppressors, which can augment the functionality of p53 by repressing the expression of its main negative regulators MDM2 and MDM4.

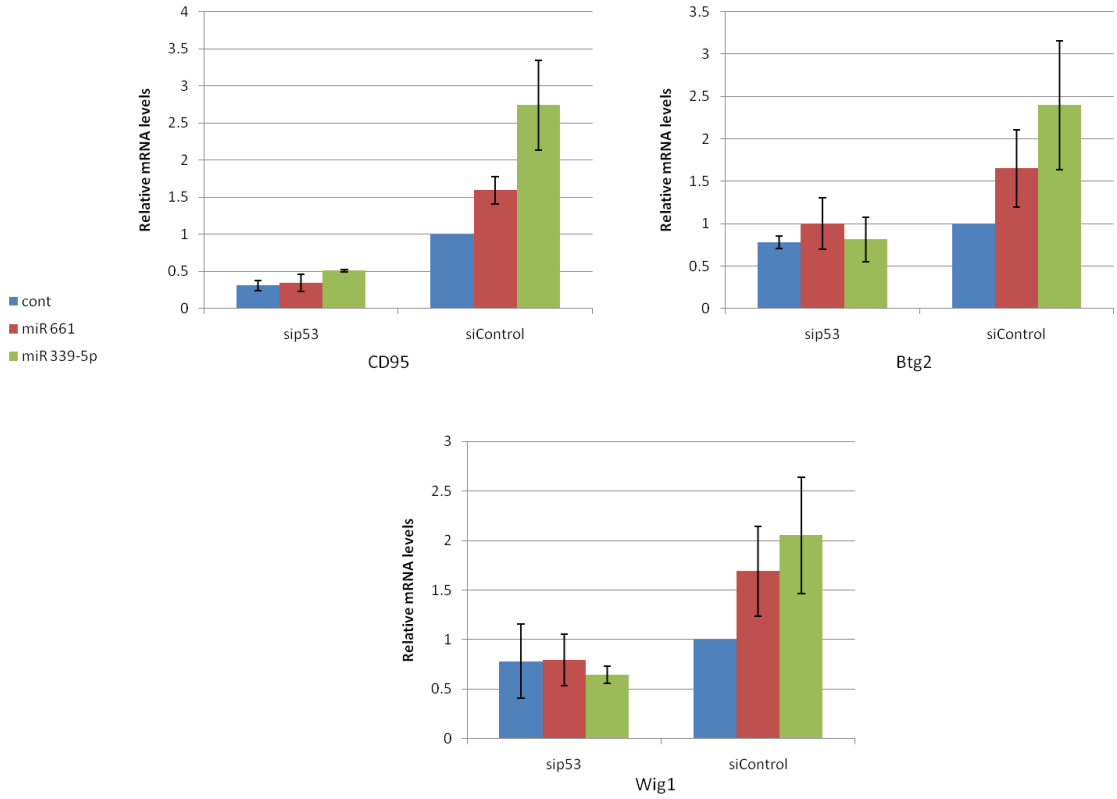

Figure 23: miR-661 and miR-339-5p upregulate various p53 target genes.

MCF7 cells were transfected with 20 nM miR-661 or miR-339-5p or miR-control and 20nM p53 siRNA or control siRNA. 48 hours later, cells were subjected to total RNA extraction, followed by qRT-PCR analysis for transcripts of the *CD95*, *Btg2* and *Wig1* genes. Values were normalized to *GAPDH* mRNA in the same sample.

miR-661 overexpression inhibits cell cycle progression and elicits apoptosis

To investigate the biological impact of the upregulation of p53 target genes by miR-661 overexpression, I next examined the effect of this miR on the cell cycle profile. This was done with and without DNA damage, where the DNA damage served to further activate the endogenous p53.

As can be seen in figure 24A, transfection of MCF7 cells with miR-661 led to a decrease in the S and G2/M phases; this is consistent with the upregulation of p21 which is known to inhibit the entry of cells from G1 into S phase and cause a growth arrest in late G1. In addition, miR-661 overexpression caused an increase in the percentage of cells with subG1 DNA content, indicative of apoptosis, consistent with the observation that miR-661 drives the upregulation of pro-apoptotic genes such as *CD95*.

When DNA damage (Doxorubicin) was introduced (figure 24B), the pro-apoptotic effect of miR-661 was further enhanced. Under those conditions, miR-661 led to a 3 fold increase in the percentage of apoptotic cells. In addition, there was also a significant effect on the cell cycle: more than 2 fold decrease in the S and G2/M phases.

In conclusion, miR-661 overexpression can promote cell cycle arrest and apoptosis, both under basal conditions, and even more so after exposure to DNA damage.

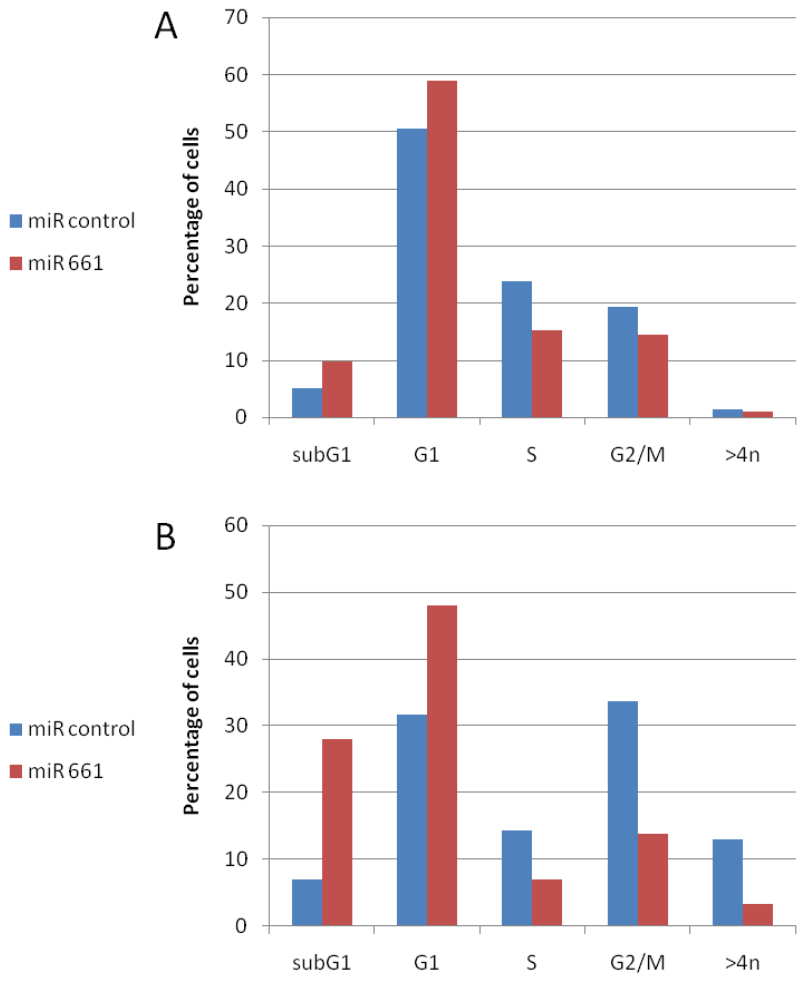

Figure 24: miR-661 overexpression elicits cell cycle arrest and apoptosis. MCF7 cells were transfected with 20nM miR-661 or miR-control for 48 hours, and then either left untreated (A) or exposed to 50 μM Etoposide for an additional 24h (B). Cells were then harvested, fixed, stained with PI and subjected to cell cycle analysis by flow cytometry.

miR-661 modulates MDM2, MDM4 and p21 protein levels in cells exposed to DNA damage

Exposure to DNA damage induces p53-dependent transcriptional upregulation of many p53 target genes. I therefore monitored the effect of miR-661 upregulation on various p53-relevant proteins in cells exposed to DNA damage. To that end, cultures were treated with Doxorubicin and the cells were then harvested at different time points and subjected to Western blot analysis. As can be seen in figure 25, the levels of p53 increased over time, followed by an increase of MDM2. However, cells overexpressing miR-661 displayed lower levels of MDM2 relative to miR control-transfected cells. As previously reported, MDM4 levels are not significantly affected by DNA damage, at least for the duration of this time course; here, too, a negative effect of miR-661 was clearly seen at all time points. The levels of p21 increased over time, as expected, but in cells transfected with miR-661 this increase was more pronounced, presumably as a result of the reduction in MDM2 and MDM4 levels that causes an upregulation of p53 functionality.

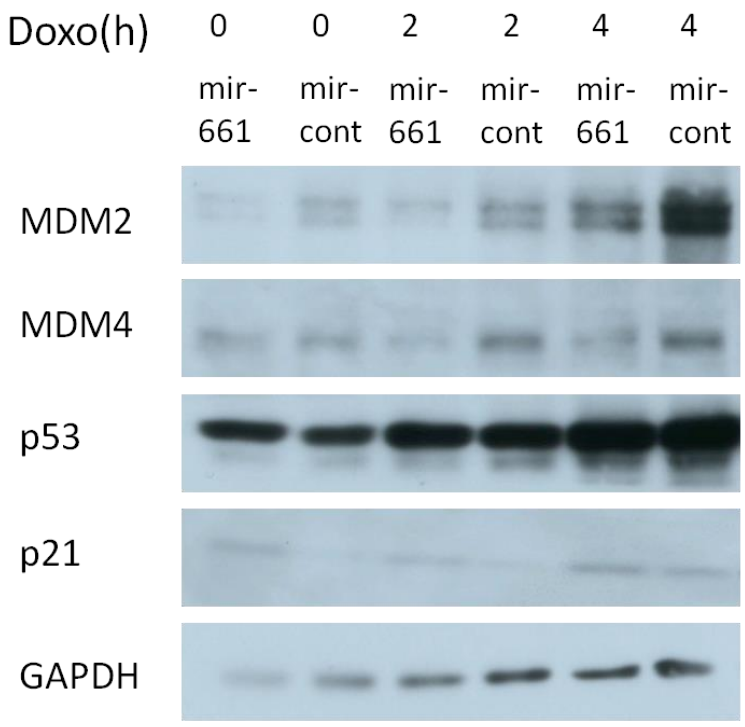

Figure 25: miR-661 modulates MDM2, MDM4 and p21 protein levels in cells exposed to DNA damage. MCF7 cells were transfected with 20 nM miR-661 or miR-control for 48 hours, followed by addition of 1μM doxorubicin (Doxo) for 0, 2 or 4 hours. Proteins were analyzed by Western blot with the indicated antibodies.

Discussion

In this work I attempted to examine the interplay between miRs and Alu sequences. Alus are primate specific, and have more than one million copies in the human genome[8]. 16% of all genes contain at least one Alu in their 3'UTR, which implies that evolutionarily new sequences have entered into a major regulatory region of genes. I specifically looked at the Alus' ability to insert new miR targets into the 3'UTR of genes, which may alter the whole regulation of the gene, and also introduce new coordinated regulatory rule into a group of genes into which the same Alu has entered. Although some of these new targets may serve as raw material for evolutionary novelty, a substantial challenge for the cell is to mask the rest from a potential distorting regulatory effect.

Here I report that the majority of miR targets within Alus are non-functional, as presented in figure 5. I show that the percentage of downregulated genes, as well as the average fold change upon miR over-expression, are significantly higher in miR binding sites located outside Alus relative to those residing within miRs. In addition I show that, on average, miR binding sites inside Alus tend to be inactive, as reflected by the observation that their effects on gene expression are not significantly greater than seen for genes that have no target at all for the tested miR. After concluding that miR binding sites within Alus are often non functional miR targets, I set out to investigate the mechanisms that allow the cell to mask the effect of those legitimate-looking miR targets. I could obtain evidence to support three such mechanisms: location, structure and RNA editing. While conserved and therefore presumably functional miR binding sites tend to concentrate near the beginning and end of the 3'UTR [5] (~250 bp), Alus often reside in the middle of the 3'UTR, which implies that Alus inserted near the ends were likely selected against during evolution. Such negative selection may be explained by a variety of scenarios. For instance, Alus inserted near the ends might have forced important miR targets to move towards the middle of the 3'UTR, where they become less effective. Another scenario, supported by my findings, is that Alus near the ends might have introduced new miR binding sites in locations where such targets are highly functional, grossly altering the conserved regulation of the gene. Additionally, it is of course conceivable that insertion of Alus near the ends of 3'UTRs may be deleterious also for reasons that are unrelated to miR function.

The second mechanism that may disregard the RNA silencing machinery to disregard miR targets within Alus is the secondary structure of such target regions, which tends to be particularly tight. Such structure renders the recognition and binding of the target by the RISC machinery energetically less favorable[77].

A third possible mechanism is RNA editing. RNA editing occurs predominantly in Alu sequences[80-84]. As such, editing is also likely to alter miR binding sites within Alus, making them dysfunctional. My analysis supports such conjecture; however, the evidence is not conclusive enough at this point, because the great majority of the miR targets in the Khan dataset contain at least one A and are thus potentially subject to editing. The number of targets that are totally A-free is therefore too small to enable a comparison between them and the A-containing sites that yields sufficient statistical power. To address this problem I intend to conduct an expression array experiment where miR-373 will be overexpressed in cells with or without simultaneous overexpression or silencing of the ADAR enzyme, which performs the editing[88]. I hope to be able to demonstrate that miR biding sites within Alus become even less functional in the presence of ADAR, wheeas A-free sites are not affected. Such proof can open new research possibilities, since ADAR levels are frequently reduced in cancer[89], raising the possibility that miR binding sites within Alus become more functional in cancer cells.
To conclude, the cell uses at least three different mechanisms to mask the possibly deleterious effects of miR binding sites within Alus. This involves a combination of evolutionary means (location), Alu-contained inherent features (structure), and an enzymatic mechanism (editing), all in order to "hide" the new miR binding sites that Alus introduce into the 3'UTR from the miR machinery. Despite the above it is intriguing that xx miR sites in the genome reside in Alus but they appear not be masked by any of the three means discussed here. A subset of xx of them are contained in the Kahan dataset and are indeed shown to function as *bona fide* miR targets. These results suggest that there is considerable potential for Alus to have introduced genomic novelty into the human genome.

Indeed in the complementary half of this study I identified miR-661 and miR339-5p as new regulators of MDM2 and MDM4 that partially act from within Alu elements . I showed that both miRs downregulate MDM2 mRNA and protein levels, while miR-661 downregulates MDM4 protein levels as well. As a result of MDM2 and MDM4 downregulation by miR-661 and miR-339-5p overexpression, p53 activity is upregulated. I showed that, at least in MCF7 cells, the protein levels of p53 are not increased upon overexpression of those miRs, but its activity does. This is reflected by an increase in the mRNA levels of p53 target genes such as *p21*, *CD95*, *Btg2* and *Wig1* after the overexpression of the miRs, an increase that was shown to be dependent on p53 since it was abolished by p53 knockdown. Moreover, miR-661 exerts a significant effect on the cell cycle: miR-661 overexpression leads to a G1 arrest and an increase in apoptosis, which can be attributed to the

upregulation of p53 target genes such as *p21* and *CD95*. Both biological effects became better evident when DNA damage was introduced to boost the impact of p53 on cell fate. Together, these results support the conjecture that miR-661 may function as a tumor suppressor miR, which is in correlation with Reddy *et al*, who showed that miR-661 downregulates MTA1 and that this miR is downregulated during tumor progression[60].

In the future it will be interesting to evaluate the effects of downregulation of endogenous miR-661 in cancer and normal cells. In addition, it will be important to apply more definitive approaches towards proving that *MDM2* and *MDM4* mRNAs are direct targets of miR-661. This might be obtained by inserting the putative miR-661 into a suitable luciferase reporter, and particularly by performing pull-down experiments to demonstrate a physical interaction between miR-661 and these two mRNAs. Identification of additional miR-661 targets is also important in order to understand in full the molecular mechanisms underlying its biological effects.

This thesis thus connects between two seemingly separate areas – Alus in the human genome and regulation by miR of the p53 pathwayParticularly interesting is my observation that most of the target sequences for miR-661 and miR-339-5p in *MDM2* and *MDM4* are contained within Alus. More specifically, miR-661 overexpression has a strong effect on MDM2 levels; of note, *MDM2* mRNA has 3 predicted targets for miR-661, all residing within Alus. Such result implies that the conclusion that targets within Alus are nonfunctional may have very interesting exceptions. It will be very interesting to find out how particular mRNAs become escape the generally applied masking of its Alu-contained miR sites. For example, do they have unusually "loose" secondary structure in the target region as compared to other Alus? Are there specific proteins that bind to those mRNA and open up the secondary structure? Or perhaps efficacy is achieved through the presence of multiple miR binding sites within Alus in the same mRNA (as is the case for both *MDM2* and *MDM4* mRNA), each of which is not functional on its own but can contribute through cooperativity of those multiple sites? Lastly, it is noteworthy that since Alus exist only in primate genomes, such regulation of targets within Alus is by definition primate-specific. This means that the relationship between miR-661 and MDM2 is evolutionarily new. Since the p53 pathway is highly conserved in mammals, it is surprising to see a regulation that is primate-specific. This may imply that, as part of the improved anti-cancer defense in relatively long-lived organisms such as primates, the p53 network has been expanded to

incorporate new regulatory components and perhaps even new regulatory principles, which extend the versatility and sensitivity of this network.

References

1 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).

2 Lu, M. *et al.* An analysis of human microRNA and disease associations. *PLoS One* **3**, e3420, doi:10.1371/journal.pone.0003420 (2008).

3 Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-866, doi:nrc1997 [pii]
10.1038/nrc1997 (2006).

4 Thomas, M., Lieberman, J. & Lal, A. Desperately seeking microRNA targets. *Nat Struct Mol Biol* **17**, 1169-1174, doi:nsmb.1921 [pii]
10.1038/nsmb.1921 (2010).

5 Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).

6 Saito, T. & Saetrom, P. MicroRNAs--targeting and target prediction. *N Biotechnol* **27**, 243-249, doi:S1871-6784(10)00372-9 [pii]
10.1016/j.nbt.2010.02.016 (2010).

7 Min, H. & Yoon, S. Got target? Computational methods for microRNA target prediction and their extension. *Exp Mol Med* **42**, 233-244, doi:emm.2010.42.032 [pii] (2010).

8 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

9 Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J. & Schmitz, J. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet* **23**, 158-161 (2007).

10 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009).

11 Cordaux, R., Hedges, D. J., Herke, S. W. & Batzer, M. A. Estimating the retrotransposition rate of human Alu elements. *Gene* **373**, 134-137, doi:S0378-1119(06)00070-9 [pii]
10.1016/j.gene.2006.01.019 (2006).

12 Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-48, doi:10.1038/ng1223
ng1223 [pii] (2003).

13 Grover, D., Mukerji, M., Bhatnagar, P., Kannan, K. & Brahmachari, S. K. Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* **20**, 813-817 (2004).

14 An, H. J., Lee, D., Lee, K. H. & Bhak, J. The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3' untranslated regions. *BMC Genomics* **5**, 97 (2004).

15 Liu, W. M., Chu, W. M., Choudary, P. V. & Schmid, C. W. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* **23**, 1758-1765, doi:5s0044 [pii] (1995).

16 Vansant, G. & Reynolds, W. F. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc Natl Acad Sci U S A* **92**, 8229-8233 (1995).

17 Polak, P. & Domany, E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* **7**, 133, doi:1471-2164-7-133 [pii]
10.1186/1471-2164-7-133 (2006).

18 Lev-Maor, G. *et al.* Intronic Alus influence alternative splicing. *PLoS Genet* **4**, e1000204, doi:10.1371/journal.pgen.1000204 (2008).
19 Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. *Genome Res* **12**, 1060-1067, doi:10.1101/gr.229302 (2002).
20 Chen, C. Y. *et al.* AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**, 451-464, doi:S0092-8674(01)00578-5 [pii] (2001).
21 Mukherjee, D. *et al.* The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J* **21**, 165-174, doi:10.1093/emboj/21.1.165 (2002).
22 Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**, 284-288, doi:nature09701 [pii] 10.1038/nature09701 (2011).
23 Chen, C., Ara, T. & Gautheret, D. Using Alu elements as polyadenylation sites: A case of retroposon exaptation. *Mol Biol Evol* **26**, 327-334, doi:msn249 [pii] 10.1093/molbev/msn249 (2009).
24 Yuan, Z. *et al.* Origin and evolution of a placental-specific microRNA family in the human genome. *BMC Evol Biol* **10**, 346, doi:1471-2148-10-346 [pii] 10.1186/1471-2148-10-346 (2010).
25 Piriyapongsa, J., Marino-Ramirez, L. & Jordan, I. K. Origin and evolution of human microRNAs from transposable elements. *Genetics* **176**, 1323-1337 (2007).
26 Bailey, J. A., Liu, G. & Eichler, E. E. An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* **73**, 823-834, doi:S0002-9297(07)63631-3 [pii] 10.1086/378594 (2003).
27 Jurka, J., Kapitonov, V. V., Kohany, O. & Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet* **8**, 241-259, doi:10.1146/annurev.genom.8.080706.092416 (2007).
28 Smalheiser, N. R. & Torvik, V. I. Mammalian microRNAs derived from genomic repeats. *Trends Genet* **21**, 322-326 (2005).
29 Smalheiser, N. R. & Torvik, V. I. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet* **22**, 532-536 (2006).
30 Lehnert, S. *et al.* Evidence for co-evolution between human microRNAs and Alu-repeats. *PLoS One* **4**, e4456, doi:10.1371/journal.pone.0004456 (2009).
31 Shalgi, R., Pilpel, Y. & Oren, M. Repression of transposable-elements - a microRNA anti-cancer defense mechanism? *Trends Genet* **26**, 253-259, doi:S0168-9525(10)00066-1 [pii] 10.1016/j.tig.2010.03.006 (2010).
32 Harms, K., Nozell, S. & Chen, X. The common and distinct target genes of the p53 family transcription factors. *Cell Mol Life Sci* **61**, 822-842, doi:10.1007/s00018-003-3304-4 (2004).
33 Green, D. R. & Chipuk, J. E. p53 and metabolism: Inside the TIGAR. *Cell* **126**, 30-32, doi:S0092-8674(06)00825-7 [pii] 10.1016/j.cell.2006.06.032 (2006).
34 Marine, J. C. & Lozano, G. Mdm2-mediated ubiquitylation: p53 and beyond. *Cell Death Differ* **17**, 93-102, doi:cdd200968 [pii] 10.1038/cdd.2009.68 (2010).
35 Sinnett, D., Richer, C., Deragon, J. M. & Labuda, D. Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J Biol Chem* **266**, 8675-8678 (1991).

36 Zuckerman, V., Wolyniec, K., Sionov, R. V., Haupt, S. & Haupt, Y. Tumour suppression by p53: the importance of apoptosis and cellular senescence. *J Pathol* **219**, 3-15, doi:10.1002/path.2584 (2009).
37 Weinberg, W. C. & Denning, M. F. P21Waf1 control of epithelial cell cycle and cell fate. *Crit Rev Oral Biol Med* **13**, 453-464 (2002).
38 Moll, U. M., Wolff, S., Speidel, D. & Deppert, W. Transcription-independent pro-apoptotic functions of p53. *Curr Opin Cell Biol* **17**, 631-636, doi:S0955-0674(05)00146-8 [pii] 10.1016/j.ceb.2005.09.007 (2005).
39 Zhang, Y., Xiong, Y. & Yarbrough, W. G. ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways. *Cell* **92**, 725-734, doi:S0092-8674(00)81401-4 [pii] (1998).
40 de Stanchina, E. *et al.* PML is a direct p53 target that modulates p53 effector functions. *Mol Cell* **13**, 523-535, doi:S1097276504000620 [pii] (2004).
41 Ferbeyre, G. *et al.* PML is induced by oncogenic ras and promotes premature senescence. *Genes Dev* **14**, 2015-2027 (2000).
42 Pearson, M. *et al.* PML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature* **406**, 207-210, doi:10.1038/35018127 (2000).
43 Perry, M. E. The regulation of the p53-mediated stress response by MDM2 and MDM4. *Cold Spring Harb Perspect Biol* **2**, a000968, doi:10.1101/cshperspect.a000968 (2010).
44 Sharp, D. A., Kratowicz, S. A., Sank, M. J. & George, D. L. Stabilization of the MDM2 oncoprotein by interaction with the structurally related MDMX protein. *J Biol Chem* **274**, 38189-38196 (1999).
45 Toledo, F. & Wahl, G. M. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat Rev Cancer* **6**, 909-923, doi:nrc2012 [pii] 10.1038/nrc2012 (2006).
46 Honda, R. & Yasuda, H. Activity of MDM2, a ubiquitin ligase, toward p53 or itself is dependent on the RING finger domain of the ligase. *Oncogene* **19**, 1473-1476, doi:10.1038/sj.onc.1203464 (2000).
47 Momand, J., Zambetti, G. P., Olson, D. C., George, D. & Levine, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237-1245, doi:0092-8674(92)90644-R [pii] (1992).
48 Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296-299, doi:10.1038/387296a0 (1997).
49 Thut, C. J., Goodrich, J. A. & Tjian, R. Repression of p53-mediated transcription by MDM2: a dual mechanism. *Genes Dev* **11**, 1974-1986 (1997).
50 Perry, M. E., Piette, J., Zawadzki, J. A., Harvey, D. & Levine, A. J. The mdm-2 gene is induced in response to UV light in a p53-dependent manner. *Proc Natl Acad Sci U S A* **90**, 11623-11627 (1993).
51 Linares, L. K., Hengstermann, A., Ciechanover, A., Muller, S. & Scheffner, M. HdmX stimulates Hdm2-mediated ubiquitination and degradation of p53. *Proc Natl Acad Sci U S A* **100**, 12009-12014, doi:10.1073/pnas.2030930100 2030930100 [pii] (2003).
52 Barboza, J. A., Iwakuma, T., Terzian, T., El-Naggar, A. K. & Lozano, G. Mdm2 and Mdm4 loss regulates distinct p53 activities. *Mol Cancer Res* **6**, 947-954, doi:6/6/947 [pii] 10.1158/1541-7786.MCR-07-2079 (2008).
53 Shvarts, A. *et al.* MDMX: a novel p53-binding protein with some functional properties of MDM2. *EMBO J* **15**, 5349-5357 (1996).

54 Marine, J. C. & Jochemsen, A. G. Mdmx as an essential regulator of p53 activity. *Biochem Biophys Res Commun* **331**, 750-760, doi:S0006-291X(05)00657-1 [pii]
10.1016/j.bbrc.2005.03.151 (2005).
55 Danovi, D. *et al.* Amplification of Mdmx (or Mdm4) directly contributes to tumor formation by inhibiting p53 tumor suppressor activity. *Mol Cell Biol* **24**, 5835-5843, doi:10.1128/MCB.24.13.5835-5843.2004
24/13/5835 [pii] (2004).
56 Migliorini, D. *et al.* Hdmx recruitment into the nucleus by Hdm2 is essential for its ability to regulate p53 stability and transactivation. *J Biol Chem* **277**, 7318-7323, doi:10.1074/jbc.M108795200
M108795200 [pii] (2002).
57 Lenos, K. & Jochemsen, A. G. Functions of MDMX in the modulation of the p53-response. *J Biomed Biotechnol* **2011**, 876173, doi:10.1155/2011/876173 (2011).
58 Toh, Y., Kuwano, H., Mori, M., Nicolson, G. L. & Sugimachi, K. Overexpression of metastasis-associated MTA1 mRNA in invasive oesophageal carcinomas. *Br J Cancer* **79**, 1723-1726, doi:10.1038/sj.bjc.6690274 (1999).
59 Jang, K. S., Paik, S. S., Chung, H., Oh, Y. H. & Kong, G. MTA1 overexpression correlates significantly with tumor grade and angiogenesis in human breast cancers. *Cancer Sci* **97**, 374-379, doi:CAS [pii]
10.1111/j.1349-7006.2006.00186.x (2006).
60 Reddy, S. D., Pakala, S. B., Ohshiro, K., Rayala, S. K. & Kumar, R. MicroRNA-661, a c/EBPalpha target, inhibits metastatic tumor antigen 1 and regulates its functions. *Cancer Res* **69**, 5639-5642, doi:0008-5472.CAN-09-0898 [pii]
10.1158/0008-5472.CAN-09-0898 (2009).
61 Bui-Nguyen, T. M. *et al.* Stimulation of inducible nitric oxide by hepatitis B virus transactivator protein HBx requires MTA1 coregulator. *J Biol Chem* **285**, 6980-6986, doi:M109.065987 [pii]
10.1074/jbc.M109.065987 (2010).
62 Su, F., Theodosis, C. N. & Schneider, R. J. Role of NF-kappaB and myc proteins in apoptosis induced by hepatitis B virus HBx protein. *J Virol* **75**, 215-225, doi:10.1128/JVI.75.1.215-225.2001 (2001).
63 Majano, P. *et al.* Hepatitis B virus X protein transactivates inducible nitric oxide synthase gene promoter through the proximal nuclear factor kappaB-binding site: evidence that cytoplasmic location of X protein is essential for gene transactivation. *Hepatology* **34**, 1218-1224, doi:S0270913901355933 [pii]
10.1053/jhep.2001.29626 (2001).
64 Yun, C. *et al.* NF-kappaB activation by hepatitis B virus X (HBx) protein shifts the cellular fate toward survival. *Cancer Lett* **184**, 97-104, doi:S0304383502001878 [pii] (2002).
65 Bui-Nguyen, T. M. *et al.* NF-kappaB signaling mediates the induction of MTA1 by hepatitis B virus transactivator protein HBx. *Oncogene* **29**, 1179-1189, doi:onc2009404 [pii]
10.1038/onc.2009.404 (2010).
66 Vetter, G. *et al.* miR-661 expression in SNAI1-induced epithelial to mesenchymal transition contributes to breast cancer cell invasion by targeting Nectin-1 and StarD10 messengers. *Oncogene* **29**, 4436-4448, doi:onc2010181 [pii]
10.1038/onc.2010.181 (2010).
67 Ueda, R. *et al.* Dicer-regulated microRNAs 222 and 339 promote resistance of cancer cells to cytotoxic T-lymphocytes by down-regulation of ICAM-1. *Proc Natl Acad Sci U S A* **106**, 10746-10751, doi:0811817106 [pii]

10.1073/pnas.0811817106 (2009).
68 Ju, X. *et al.* Differential microRNA expression in childhood B-cell precursor acute lymphoblastic leukemia. *Pediatr Hematol Oncol* **26**, 1-10, doi:908622265 [pii] 10.1080/08880010802378338 (2009).
69 Wu, Z. S. *et al.* MiR-339-5p inhibits breast cancer cell migration and invasion in vitro and may be a potential biomarker for breast cancer prognosis. *BMC Cancer* **10**, 542, doi:1471-2407-10-542 [pii] 10.1186/1471-2407-10-542 (2010).
70 Ichi, S. *et al.* Folic acid remodels chromatin on Hes1 and Neurog2 promoters during caudal neural tube development. *J Biol Chem* **285**, 36922-36932, doi:M110.126714 [pii] 10.1074/jbc.M110.126714 (2010).
71 Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* **39**, D876-882, doi:gkq963 [pii] 10.1093/nar/gkq963 (2011).
72 Khan, A. A. *et al.* Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* **27**, 549-555 (2009).
73 Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20 (2005).
74 Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**, 25-36, doi:S0092867403002319 [pii] (2003).
75 Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev* **17**, 438-442, doi:10.1101/gad.1064703 (2003).
76 Dahary, D., Shalgi, R. & Pilpel, Y. CpG Islands as a putative source for animal miRNAs: evolutionary and functional implications. *Mol Biol Evol* **28**, 1545-1551, doi:msq315 [pii] 10.1093/molbev/msq315 (2011).
77 Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**, 1278-1284 (2007).
78 Labuda, D. & Striker, G. Sequence conservation in Alu evolution. *Nucleic Acids Res* **17**, 2477-2491 (1989).
79 Okada, N. Transfer RNA-like structure of the human Alu family: implications of its generation mechanism and possible functions. *J Mol Evol* **31**, 500-510 (1990).
80 Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* **2**, e391 (2004).
81 Blow, M., Futreal, P. A., Wooster, R. & Stratton, M. R. A survey of RNA editing in human brain. *Genome Res* **14**, 2379-2387 (2004).
82 Kim, D. D. *et al.* Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res* **14**, 1719-1725 (2004).
83 Levanon, E. Y. *et al.* Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**, 1001-1005 (2004).
84 Morse, D. P., Aruscavage, P. J. & Bass, B. L. RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. *Proc Natl Acad Sci U S A* **99**, 7906-7911 (2002).
85 Ullu, E. & Tschudi, C. Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171-172 (1984).

86 Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**, D501-504 (2005).

87 Jelinek, W. R. & Schmid, C. W. Repetitive sequences in eukaryotic DNA and their expression. *Annu Rev Biochem* **51**, 813-844 (1982).

88 Bass, B. L. RNA editing and hypermutation by adenosine deamination. *Trends Biochem Sci* **22**, 157-162, doi:S0968000497010359 [pii] (1997).

89 Paz, N. *et al.* Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res* **17**, 1586-1595, doi:gr.6493107 [pii]
10.1101/gr.6493107 (2007).

90 Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-3.0.
1996-2010 <http://www.repeatmasker.org>