מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

מאת
יונית הופמן

מיתוג של מיקרו-רנ"א לאורך ציר זמן פיסיולוגי ואבולוציוני

Switching microRNAs ON and OFF on physiological and evolutionary time scales

מנחים:
פרופ' יצחק פלפל
פרופ' משה אורן

תמוז 2016

Thesis for the degree
Doctor of Philosophy

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

By
Yonit Hoffman

Advisors:
Prof. Yitzhak Pilpel
Prof. Moshe Oren

August 2016

Table of Contents

1. Abstract

microRNAs (miRNAs) are short non-coding RNAs that target mRNA molecules and down-regulate the expression of the corresponding proteins. They do so primarily by binding to a 7-mer in the 3'UTR of the mRNA. As such, miRNAs are key regulators of cellular pathways. There are many potential miRNA binding sites in each gene's 3'UTR; however, the majority of them are non-functional. Various features within and around the binding site can affect its functionality, and my research was focused on two main such features: Alu elements and alternative polyadenylation.

Alu sequences are widely spread, primate-specific retrotransposons, which contain many potential binding sites for highly conserved miRNAs, and therefore theoretically have the ability to spread miRNA regulation all over the genome. They could provide an evolutionary opportunity to the genome for conferring new primate-specific regulation via a single insertion of an Alu element, but at the same time Alus are also able to enter new genes and alter and even distort their regulation. I found that on a global perspective, miRNA binding sites within Alus are significantly less responsive to the miRNA machinery than binding sites residing outside Alu elements, thus suggesting that the miRNA machinery can avoid binding to them in most cases. In contrast, I did find a particular case that represents the opposite: a primate-specific miRNA – miR-661, which targets the p53 pathway via two of its important regulators, Mdm2 and Mdm4, presumably via binding sites within Alus residing in *Mdm2* and *Mdm4* mRNA. I showed that miR-661 has an effect on the cell cycle, and high levels of this miRNA are correlated with better outcome for breast cancer patients. Therefore, although the global effect of miRNA binding sites within Alus presents a challenge to the genome, and hence they are often being "masked" and non-responsive to the corresponding miRNAs, in specific cases such as in the p53 network they can actually contribute new primate-specific regulation.

miRNA regulation occurs mainly via the 3'UTR of their target mRNAs, which in many cases can harbor multiple potential polyadenylation sites. Besides the canonical site that determines the end of the 3'UTR in most mRNA molecules, there are other sites (alternative polyadenylation sites – APA) that are functional in specific cases and cell lines. More than 50% of the mRNAs have more than a single polyadenylation site. miRNA binding sites tend to be more responsive to the miRNA machinery when they are positioned near the beginning or the end of the 3'UTR. Therefore, an

interesting question is what happens when an APA cleavage occurs, and a new set of miRNA binding sites are now positioned near the end of the 3'UTR? I found that the area immediately upstream to the APA is enriched with conserved miRNA binding sites, which may become functionalized when the APA site is used. This phenomenon occurs in several human cell lines that I tested, and in the mouse as well. Interestingly pro-differentiation genes tend to be enriched for conserved miRNA binding sites prior to their APA sites, which may constitute a mechanism for downregulation of such genes during proliferation and cancer. Moreover, miRNAs that I and others found to be upregulated in proliferation and cancer, particularly the miR-17-92 cluster, are enriched for conserved binding sites prior to their APA sites, and relatively depleted of such sites near the end of their full-length 3'UTR.

The tendency of miRNA binding sites within Alus to be almost non-functional is a feature of the miRNA binding site itself, and in an evolutionary time scale these binding sites were shut off, or were not able to be turned on. On the other hand, APA events that can functionalize miRNA binding sites represent another layer of regulation – a gene-specific behavior that implies whether a binding site will work or not. This layer is specifically unique, as it can be changed during a physiological time scale, where for the same gene a binding site might be functional or non-functional depending on an APA event.

1. תקציר

מיקרורנ"א הינן מולקולות קצרות של רנ"א שאינן מקודדות לחלבון, ומבקרות באופן שלילי את ביטויים של חלבוני מטרה. הן עושות זאת על ידי קישור של כ – 7 בסיסים לאזור ה 3'UTR של המרנ"א. מיקרורנ"א הינם בקרים חשובים במסלולים תאיים. ישנם אתרים רבים בעלי פוטנציאל קישור למיקרורנ"א בכל מרנ"א, אולם רובם המכריע אינו שמיש. מאפיינים שונים של אתר הקישור וסביבתו משפיעים על יכולתו לשמש ככזה, והמחקר שלי מתמקד בשני מאפיינים מרכזיים: רצפי אלו (Alu) ופוליאדנילציה חלופית.

רצפי אלו הם רצפים נפוצים מאד בגנום הפרימטי, הם מסוג רטרו-טרנספוזונים והם מכילים אתרי קישור פוטנציאלים רבים למיקרורנ"א, ולכן לפחות באופן תיאורטי מסוגלים להפיץ את בקרת המיקרורנ"א בכל הגנום. הם עשויים להוות הזדמנות אבולוציונית לגנום ליצירת בקרה חדשה פרימטית על ידי הכנסה יחידה של רצף אלו, אולם בו בזמן הם עלולים לקפוץ לגנים חדשים ולשנות את כל מערך הבקרה הקיים. אני מצאתי שבאופן גלובאלי, אתרי קישור למיקרורנ"א בתוך רצפי אלו כמעט ואינם שמישים בהשוואה לרצפי קישור מחוץ לאלו, והמיקרורנ"א נמנעים מלהיקשר אליהם במרבית המקרים. לעומת זאת, מצאתי דוגמא יוצאת דופן של מיקרורנ"א ספציפי לפרימטים – מיר-661 – שמבקר את מסלול העברת האותות p53 על ידי בקרה שלילית לשני חלבונים חשובים במסלול – Mdm2 ו Mdm4, ככל הנראה על ידי אתרי קישור בתוך רצפי אלו. הראיתי שלמיר-661 יש אפקט על מסלול התא, ורמות גבוהות שלו נמצאות בקורלציה עם פרוגנוזה טובה יותר בחולות סרטן השד. לכן, למרות שההשפעה הגלובלית של אתרי קישור למיקרורנ"א בתוך רצפי אלו מהווה אתגר לגנום ולכן ממוסכת על ידו ומיקרורנ"א נמנעים מלהיקשר לאתרים אלו, ישנם מקרים מסוימים, כמו ברשת של p53, בהם אתרים אלו תורמים לבקרה שהינה ייחודית לפרימטים.

בקרה של מיקרורנ"א על מרנ"א קורית בעיקר דרך קישור לאזור ה 3'UTR, אשר במקרים רבים יכול להכיל מספר אתרי סיום פוטנציאלים (אתרי פוליאדנילציה). מלבד אתר הפוליאדנילציה הקנוני אשר מהווה את סיום המרנ"א במרבית המולקולות, ישנם אתרי פוליאדנילציה חלופיים שהינם שמישים במצבים מסוימים או בתאים שונים. למעלה ממחצית המרנ"א יש יותר מאתר פוליאדנילציה אחד. אתרי קישור למיקרורנ"א נוטים להיות שמישים יותר כאשר הם ממוקמים קרוב להתחלה ולסוף של ה 3'UTR, ולכן שאלה מעניינת היא מה קורה כאשר ישנה פוליאדנילציה חלופית, וקבוצה חדשה של אתרי קישור למיקרורנ"א כעת ממוקמת קרוב לסוף החדש של ה 3'UTR? אני מצאתי שהאזור שנמצא בדיוק לפני אתר הפוליאדנילציה החלופית מועשר באתרי קישור שמורים למיקרורנ"א, אשר יכולים להיות משופעלים כאשר אתר הפוליאדנילציה החלופית נמצא בשימוש. תופעה זו קורית בסוגים שונים של תאים הומניים שנבדקו וכן בתאי עכבר. גנים שהינם מעודדי דיפרנציאציה נוטים להיות מועשרים יותר לאתרי קישור שמורים למיקרורנ"א באזורים שלפני אתרי הפוליאדנילציה החלופית, מה שעשוי להעיד על מנגנון להורדת רמות ביטוי גנים אלו בזמן פרוליפרציה וסרטן. יתרה מכך, אתרי קישור למיקרורנ"א שנמצאו על

ידינו ועל ידי אחרים כבעלי ביטוי גבוהה בזמן פרוליפרציה וסרטן (במיוחד משפחת המיקרורנ"א מיר--17 92), מועשרים לפני אתרי פוליאדנילציה חלופית, ובאופן יחסי, למיקרורנ"א אלו יש פחות אתרי קישור שמורים ליד הקצה הארוך של ה 3'UTR.

הנטייה של אתר קישור למיקרורנ"א בתוך רצף אלו להיות לא שמיש הינה מאפיין של אתר הקישור עצמו, ובציר זמן אבולוציוני אתרים אלו הושתקו, או שלא ניתנה להם האפשרות להפוך לשמישים. מנגד, אירועי פוליאדנילציה חלופית שעשויים להשמיש אתרי קישור למיקרורנ"א מהווים שכבת בקרה חדשה – התנהגות הגן עצמו קובעת האם אתר הקישור למיקרורנ"א יהיה שמיש או לא. שכבה זו של בקרה הינה ייחודית בכך שהיא יכולה להשתנות בציר זמן פיזיולוגי של התא, כאשר עבור אותו גן אתר קישור למיקרורנ"א יכול להיות שמיש או מושתק לפי אירוע של פוליאדנילציה חלופית.

2. Introduction

2.1 microRNAs

microRNAs (miRNAs) are small non coding RNAs, ~22 nt long[1], which regulate gene expression through interaction with their mRNA targets. mRNAs recognized by a particular miRNA usually have one or several binding sites in their 3' untranslated region (UTR) that are complementary to bases 2-8 (called "seed") in the mature miRNA 5' end[2]. Although each miRNA has potentially hundreds of binding sites in at least dozens of mRNAs, it is clear that only a minority of those sites are functional, while the majority are not[3]. In addition, the functionality of a given binding site may be influenced by the identity of the cell and the physiological conditions[4]. Various features affect the functionality of a potential miRNA binding site; for example, the position along the 3'UTR[3] and the structure of the RNA around the binding site[5,6]. Other features that contribute to the sites' functionality are thermodynamic stability of seed pairing, nucleotides around the site, AU content near the site, other base pairing regions with the miRNA 3'end, and open reading frame (ORF) and 3'UTR lengths[7]. There exist numerous target prediction algorithms which use different parameters[8]. Still, it is not clear how the cell and specifically the miRNA machinery distinguish between the plethoras of potential binding sites and identify the genuine ones.

miRNAs are known to participate in many important cellular and developmental processes, and in the development of some diseases[9,10]. For example, miRNAs are involved in differentiation processes, such as in human preimplantation development[11]. They are also deeply involved in proliferation and contribute to cancerous processes[9,10], for example, in the development and progression of gastric cancer[12], colorectal cancer[13] and many more.

The processing of miRNAs involves different factors in different parts of the cell. The long miRNA primary transcript (pri-miRNA) is transcribed by RNA Polymerase II[14] and is cleaved to a pre-miRNA of 60-70 nucleotides by the nuclear RNAse III Drosha[1]. The cleavage by Drosha is near the base of the primary stem-loop of the miRNA, in both strands of the stem[1]. The next step is the export of the pre-miRNA to the cytoplasm by Ran-GTP and the export receptor Exportin-5. In the cytoplasm the pre-miRNA is cleaved to produce the mature miRNA by Dicer, which 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. Dicer cuts the pre-miRNA two helical turns from the base of the stem-loop[1]. The last step is the loading of the mature miRNA onto the RISC complex[1]. The RISC complex is composed of Ago proteins, which consist of four domains with the ability to anchor the RNA and bind the 3' end of the guide strand[15]. The first step in the RISC assembly of the miRNA is the loading of the miRNA duplex into Ago, a complex that is called the "pre-RISC". At the next step, the duplex is opened at one end by Ago, and then separated, and the passenger strand is discarded from the complex[15]. The RISC-miRNA complex is then directed to a specific subset of mRNA molecules, thus inhibiting their translation or targeting them for cleavage and degradation[1].

The decision regarding the choice of which mechanism the miRNA will apply on the mRNA for its arrest (degradation or translational block) is not clear. At first it was suggested that the degree of complementarity between the miRNA and the mRNA affects this decision, however multiple evidence contradicted this notion. Later it was suggested that the default mechanism is by translational arrest, and when perfect complementary exists, the miRNA may also lead to mRNA cleavage[16]. New data from ribosome profiling experiments suggest that the major effect of miRNA regulation is through RNA decay, while 11-16% of miRNAs effect is through translation arrest[14].

The degradation of the mRNA by the miRNA occurs because of deadenylation, decapping and exonucleolytic digestion of the mRNA, and requires Ago, GW182, and the cellular decapping and deadenylation machinery[16]. There are Ago proteins that can cleave the target mRNA by their RNase H-line PIWI domain[14]. When the miRNA leads to mRNA cleavage, the cut site it determined relative to the miRNA residues (between the nucleotides that bind residues 10 and 11 of the miRNA), and after the cleavage the miRNA can guide the RISC complex to other targets[1]. However, the miRNA-RISC complex can also repress the translation of the mRNA without its degradation, by slowing or stalling the ribosomes translating the mRNA[1]. This repression can be in the translation initiation or in a subsequent stage[16]. A new model, which was validated in zebrafish, suggests that initially miRNAs repress the translation of the mRNA and later they lead to mRNA decay, perhaps due to the block in translation and not because of the AGO proteins or the miRNAs themselves. The timing of decay after translation arrest depends on the context. If the cells have

robust surveillance mechanism, the decay will occur quickly after the translational block and vice versa[14].

2.2 Alu sequences

Repetitive elements are very widely spread in primate genomes[17]. Most prominent is the case of the Short INterspersed Elements (SINEs), and in particular the Alu repeats (Alus), which are present in more than a million copies in the human genome[18]. Most Alus were inserted ~40 million years ago in the primate lineage[19]. There is an Alu insertion every 20 births[20]. If two human genomes are compared, there are about 800 polymorphic Alu elements between them[21]. Most human genes (75%) contain an Alu sequence, in most cases in their introns[22], and 5% of the 3'UTR, on average, is comprised of Alus[23]. The density of Alus within genes is higher than in intergenic regions. An exceptional situation is observed in chromosomes 19 and 22, which contain miRNAs surrounded by Alus[22]. Transcripts of Alus are present at low level under normal conditions, but their levels increase when cells are exposed to a variety of stresses[24].

The structure of the Alu is a dimer formed by the fusion of two monomers derived from the 7SL RNA gene[25]. The two dimers form the body of the Alu, which is about 280 bases, and between them there is a short A-rich region. In the 3'end of the Alu there is also an A-rich sequence, longer than in the body, which is required for the Alu's amplification. Alus contain an internal promoter for RNA Polymerase III, with no termination signal[21].

Alus are transposable elements, which use the retrotransposition molecular machinery of the LINE TE-L1s to integrate into the host genome[18]. RNA Polymerase III transcribes the Alus, but each transcript is unique since it can accumulate mutations and differ in length and sequence of the 3'end, because there is no termination signal for the polymerase. The Alu RNA is collected into ribonucleoprotein particles. The process involves polyA binding protein (PABP), a SRP9/14 heterodimer, and other unidentified proteins, which enable the Alu to associate with the ribosome. In the ribosome, the Alu binds the ORF2 protein from the L1 elements, and then it can copy itself to a new genomic site by reverse transcription. Unlike in the process of L1 retrotransposition, Alus require only the ORF2 protein and not the ORF1 protein for their insertion into new genomic locations. This might be one of the reasons that Alu insertions are much more common in the genome[21].

Not all Alus can jump and be inserted into new locations. There are multiple factors that contribute to this ability: the sequence of the Alu, which influences its transcription, the length of the Alu and its 3' end, which is unique to each Alu, epigenetic regulation of the transcription, the length of the polyA-tail, and the sequence of the body of the Alu. The body is responsible for binding relevant proteins and for the structure of the Alu[21].

In the past, Alus were considered "junk DNA". However, recent studies show the opposite, namely that Alus have a diversity of functionalities in the promoter, the coding sequence and the UTRs of genes. One of the functions that Alus acquired over the years is the incorporation of regulatory sites. Some Alus within gene promotors (Alu classes III and IV) contain a transcription regulatory site (Retinoic Acid Response Element, RARE)[26]. During primate evolution, there was a negative selection against Alus that had been inserted near genes where a RARE site was deleterious. However, in some cases there was insertion of Alus upstream to specific genes for which Retinoic Acid inducibility was an advantage, and those Alus remained in the genome and are recognizable today, such as in the K18 gene[26]. Another regulatory feature of Alus within promoters was found by Polac et al. They showed that there are promoters that are enriched with Alus that contain multiple binding sites for transcription factors. Many of these factors are associated with early markers of development, and it was suggested that Alus were spread into promoters to insert transcription factor binding sites during evolution[27]. Alus can also contribute to splicing. Specific mutations in the Alu can create new weak splice sites, which lead to the insertion of the Alu into the protein in a process termed "Alu exonization"[28]. Examples of genes that undergo this "Alu exonization" process are a variant of the biliary glycoprotein that contains an Alu fragment, and the human decay-acceleration factor. In the human decay acceleration factor, 10% of its mRNA molecules contain an Alu[29].

In the 3'UTR, Alus were also found to have functionality. Alus in the 3'UTR can create AU rich elements – AREs; in fact ~40% of the AREs are associated with Alus[23]. The function of AREs is to destabilize the mRNA through the nuclear exosome pathway[30,31]. Another function of Alus within 3'UTR is also connected to stability of the mRNA. Staufen 1 (STAU1) mediates mRNA decay; imperfect binding of an Alu element in the 3'UTR of a SMD target (a target of STAU1) and another Alu element in a cytoplasmic, polyadenylated long non-coding RNA (lncRNA), can form

a binding site for STAU1, which affects the mRNA stability[32]. Alus can also create a polyadenylation site (PAS): Alus contain specific hexamers that represent PASs without or with one base mutation. About 107 Alus contain such active PAS[33].

The connection between miRNAs and Alus is well established, and starts at the birth of some miRNAs. About 10-20% of the miRNAs originate from transposable elements[34,35]. In particular, there is a cluster of Alus and miRNAs called C19MC, where there are many miRNAs surrounded by Alu sequences. These Alu-miRNA regions are embedded in repetitive cassettes[36]. The transcription of the miRNAs in this cluster is a consequence of the transcription of the Alus. Alu sequences contain an internal promoter for RNA polymerase III[37], without a termination sign. Therefore when Pol III transcribes the Alus in the cluster, it also transcribes the miRNAs[38]. This claim was challenged more recently, when it was shown that Pol II transcribes the miRNAs of this cluster[39]. The duplication of these Alu-miRNAs cassettes was suggested to occur because of the homology sites of Alus[40]. Several miRNAs were described to derive from genomic repeats, and as such have binding site sequences within these repeats. Many mRNAs contain Alus in their 3'UTR, and these Alus contain predicted binding sites for dozens of miRNAs[41]. Besides the known capacity of miRNAs to downregulate mRNAs, it was proposed that they also have a role in maintaining genomic stability by the repression of transposable elements[42]. Furthermore, in another study it was shown that there are some miRNAs with more than 1000 binding sites per megabase in Alus, and the majority of these miRNAs are from the C19MC cluster. miRNAs from this cluster are significantly expressed in the testis, and one possibility is that these miRNAs protect against Alu transposition[40].

Such massive presence of Alu elements, which are perceived as predominantly selfish DNA, may represent a substantial potential informational load on the genome. Accordingly, the retrotransposition of Alus may contribute to human disease, including a diversity of cancers[43]. In fact, Alus do not only contribute to genetic differences between individuals, but can contribute to genetic instability during the life of the individual, since they can be active in somatic tissues, and not only in the germ line. Such instability can lead to cancer and age-related degenerative diseases[21].

At the same time, the spread of genetic material may also represent an opportunity to introduce evolutionary novelty into the genome, as described above. Along with these negative and positive potential contributions to cellular and organismal fitness, it is conceivable that the spread of many of the retro-elements was restricted

evolutionarily so that most of the current elements are largely benign. Possibly, the insertion of retro-elements into mRNAs was not random, but was affected by features that minimize their impact on functional elements in the genome. In particular, as Alus are very abundant in the human genome, the potential effect of miRNA binding sites within them may be substantial and affect crucial cellular processes.

2.3 The p53 network

p53 is a transcription factor that responds to diverse stresses such as DNA damage, overexpressed oncogenes and various metabolic limitations[44,45]. p53 regulates the expression of a diverse group of genes that can promote cell cycle arrest, senescence, apoptosis, metabolism alteration and DNA repair[45]. By regulating these genes, p53 prevents the proliferation of genetically compromised cells[46]. 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[47].

Senescence can be triggered by oncogene activity or DNA damage. Many oncogenes such as RAS, E2F, RUNX1 and more trigger p53-induced senescence[48]. Some of them involve the DNA damage response (DDR), while others induce p53-dependent senescence without DNA damage[48]. DNA damage such as radiation, chemotherapeutic drugs or telomerase dysfunction, drive senescence primarily via the p53-p21 pathway[48]. p21 is important for DNA damage-induced senescence as well as for transient growth arrest[48]. 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[49]. Specifically, p21 inhibits cdk2 and cdk4, which are required for the progression of the cell cycle from G1 to S phase[50].

p53 controls cell cycle progression also by regulating the G2/M checkpoint, which is very important for preventing segregation of damaged chromosomes[50]. This checkpoint can be inhibited by controlling Cdc2-cyclinB activity[50]. 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[50].

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[51]. Another activator is the tumor suppressor PML (promyelocytic leukemia); expression of PML is also regulated by p53, creating a positive feedback loop between them[52]. PML stabilizes p53 through increasing its acetylation[53,54].

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 its target genes that participate in this process, and by modulating the process directly[50]. For example, p53 activates DDB2, which is important for DNA repair after UV radiation[55,56]. p53 can be activated by double stranded DNA breaks or by DNA replication arrest, via the ATM/ATR pathway. These processes lead to phosphorylation of p53 either by ATM and ATR themselves or by CHK2 and CHK1 (which are phosphorylated by ATM and ATR), all leading to activation of p53. In addition, ATM and ATR phosphorylate Mdm2, the major negative regulator of p53, which leads to its degradation[57].

Under severe stress conditions, p53 can trigger apoptosis, which has a key role in tumor suppression. The induction of apoptosis by p53 is via its target genes participating in the intrinsic and extrinsic apoptotic pathways[48]. For example, Bax, Bid, Puma and Noxa participate in the intrinsic pathway, and Fas (CD95) and DR5 death receptors, the death ligand TNFSF10, the Fas ligand TNFSF6 and caspase 8 participate in the extrinsic process[48]. The mitochondrial pathway to apoptosis is mediated by mitochondrial outer membrane permeabilization, which is activated by proteins encoded by p53 transcriptional target genes, such as Bak, Bax and Puma[58].

p53 is also induced by nutrient deprivation and metabolic imbalance, and can be involved in regulation of metabolism without causing cell-cycle arrest or apoptosis. For example, AMPK (AMP activated protein kinase) phosphorylates p53 and activates it. p53 is also involved in maintaining cellular and metabolic homeostasis under conditions of deprivation of nutrients. As examples, p53 activates Lipin1 (which has a role in regulating genes involved in fatty acid oxidation) and malonyl-CoA. p53 can also inhibit the Warburg effect (reduced oxidative phosphorylation and enhanced aerobic glycolysis), which is associated with oncogenic transformation, by upregulating p53 target genes TIGAR and SCO2 (synthesis of cytochrome oxidase 2), and inhibiting other genes such as glucose transporters[59].

The regulation of p53 is very important. One mode of regulation is at the post transcriptional modification level. p53 can be phosphorylated, and therefore

stabilized and activated, by different kinases such as ATM, ATR, Chk1, Chk2 and others. p53 can also be acetylated by histone acetyltransferase CBP/p300 in response to stress, which is important for its functionality. Different regions and timing of acetylation in p53 can lead to different cell fates. Other modifications such as methylation, sumoylation, and neddylation are also important for p53 functionality, and the combination of all these modifications can serve as a barcode for determining the specific p53 response[60].

Another layer of regulation is by protein-protein interactions, which can also mediate the p53 response to stress. Interaction of the proapoptotic factors of the ASPP family (ASPP1 and ASPP2) with p53 causes enhancement of the proapoptotic function of p53 by leading it to selectively bind to proapoptotic target genes such as Puma, Bax and others. On the other hand, another member of this family – iASSP, causes the opposite effect by selectively preventing the transcriptional activity of p53 bound to proapoptotic promoters. Another protein that causes specific promoter binding of p53 is 53BP1, which leads to cell cycle arrest, and by itself has a role in DNA repair[60].

p53 activation requires not only post transcriptional modification or cofactors involvement, but also release of p53 from its two main negative regulators – Mdm2 and Mdm4[60]. Mdm2 and Mdm4 are structurally related proteins that serve as major negative regulators of p53[61-63]. They both contain an amino-terminal p53-binding domain, a central acidic domain, and a carboxy-terminal RING finger[61,63]. Through the carboxy-terminal RING finger, Mdm2 and Mdm4 can form heterodimers, which are important for their functionality[64]. Mdm2 and Mdm4 are overexpressed in several cancer types (colorectal, esophageal, breast cancer and others), mostly with wild type p53[65]. Both Mdm2 and Mdm4 can bind to the transactivation domain of p53 and inhibit its transcriptional activity by physically blocking its interaction with components of the transcriptional machinery. In addition, Mdm2 is part of a family of ring finger containing proteins[46] and is an E3 ubiquitin ligase that can drive polyubiquitylation and subsequent proteasomal degradation of p53[61,63]. Its catalytic activity requires the ring finger for p53 ubiquitylation and for autoubiquitylation[66]. Notably, the *Mdm2* gene is a positive transcriptional target of p53[67,68]. It contains two p53 binding sites[69], underpinning a negative feedback loop that tunes down cellular p53 activity. The *Mdm4* gene does not contain such binding sites for p53, and is not regulated by p53 directly[70]. Transcription of *Mdm4* can be stimulated by

the MEK/ERK pathway or by oncogenic K-RAS expression[71]. Although Mdm4 alone has no measurable E3 activity towards p53, the Mdm2-Mdm4 hetero-oligomer is a more efficient p53 E3 ligase than Mdm2 alone, and thus Mdm4 acts as an Mdm2-dependent enhancer of p53 degradation[72]. Mdm4 can also inhibit the Mdm2 E3 ligase activity, depending on the circumstances[73]. Mdm2 loss is almost invariably lethal[47], and Mdm4 knockout is lethal for mice[60]. Thus, both Mdm2 and Mdm4 are critical in repressing p53[60]. The interaction of Mdm4 with the p53 transactivation domain can lead to reduced post-transcriptional modifications on p53, which are needed for its function[74]. For example, it can prevent the acetylation of p53 by limiting the access of the histone acetyl transferase p300 to p53[75].

The release of p53 from Mdm2 and Mdm4 is very important, and can occur by Mdm2 itself or by other factors. Mdm2 can direct its ubiquitin ligase activity towards Mdm4, for example when there is DNA damage. This causes a reduction in Mdm4 protein, leading to p53 activation[71]. When there is nucleolar disruption, ribosomal proteins are released, interact with Mdm2, and inhibit its ubiquitylation activity towards p53[71].

Inhibition of Mdm2 can lead to increased p53 functionality. Inhibition can be achieved by directly blocking the interaction of Mdm2 with p53, as in the case of the small molecule Nutlin-3a. However, Nutlin-3a was found to have poor bioavailability, and a new generation of Mdm2 small molecule inhibitors was developed. Mdm2 inhibitors can also activate non-p53 dependent pathways in cancer, for example E2F1-mediated apoptosis[76]. Mdm2 has p53-independent activities, which promote cell transformation, invasion and metastasis. Mdm2 can ubiquitinate targets besides p53, regulate the stability of specific mRNAs, and stimulate the activity of transcription factors. All these activities lead to interference with cell cycle control, apoptosis and other effects. For example, Mdm2 can directly bind p21 to negatively regulate it and lead to cell proliferation[65].

One proposed mechanism regarding the interplay between release and activation of p53, relies on the assumption that anti-repression and release (from Mdm2 and Mdm4, for example) are sufficient for the activation of target genes that are highly responsive to p53, leading to cell cycle arrest. However, for apoptosis and some other p53 induction outcomes, a more complex activation is needed[60].

2.4 miR-661

miR-661 is a human miRNA, without an ortholog in mouse. There are a few established targets of miR-661 in humans: MTA1, TUSC2, INPP5J and hTERT[77-80].

miR-661 was proposed as a tumor suppressor miRNA by targeting MTA1. MTA1 is an oncogene, upregulated in different cancers, and it is suggested as a predictor of aggressiveness of the tumor, as it is associated with tumor angiogenesis and higher tumor grade[81]. It was shown to be a target of miR-661, as its levels were downregulated when miR-661 was overexpressed. c/EBPalpha contributes to miR-661 functionality towards MTA1: c/EBPalpha binds miR-661 and upregulates it, which further downregulates the levels of MTA1. There is an anti-correlation between the levels of MTA1 and miR-661 in a breast cancer progressive isogenic model as the tumors progress (MTA1 increases and miR-661 decreases)[77]. Another study showed that miR-661 is downregulated by HBx, to upregulate MTA1, and there is a correlation between the levels of MTA1 and HBx[82]. HBx upregulates expression of iNOS (inducible nitric-oxide synthase)[83,84], and is an activator of NF-κB signaling[85]. miR-661 regulates the iNOS pathway by regulating MTA1[86].

In glioma, miR-661 was found to specifically target hTERT, and therefore inhibit cell proliferation, migration and invasion[78].

In contrast, miR-661 was shown to directly target INPP5J (inositol polyphosphate-5-phosphatase J), thereby promoting the proliferation of ovarian cancer cells. INPP5J is a negative regulator of PI3K/Akt signaling (a survival pathway activated in cancer), so miR-661 promotes the elevated activity of the PI3K/Akt pathway[79].

TUSC2 is a tumor suppressor, which is a potent proapoptotic factor. It was shown to be a target of miR-661. It also has a pseudogene with 89% homology to the *TUSC2* 3'UTR, called *TUSC2* pseudogene (*TUSC2P*). Both *TUSC2* and *TUSC2P* contain a binding site for miR-661. When both *TUSC2P* and *TUSC2* 3'UTRs were exogenously overexpressed in breast carcinoma cells, there was a decrease in cell proliferation, survival, migration, invasion and colony formation, and increased tumor cell death. This was via interaction with endogenously expressed miRNAs, including miR-661, to arrest their function and increase TUSC2 protein translation[80].

2.5 Alternative polyadenylation and 3′UTR shortening

The processing of the 3' end of the mRNA is highly important and is required for the stability of the mRNA, its nuclear export and efficient translation. It involves two steps: endonucleolytic cleavage, and addition of a poly(A) tail[87]. The 3' end processing complex contains more than 80 proteins. There are also more than 50 proteins that control the interplay with other processes[88].

The addition of the polyA requires polyA polymerase (PAP), cleavage and polyadenylation specificity factor (CPSF), and polyA binding protein (PABP). The cleavage reaction involves many components, including cleavage factor I and II, CPSF, cleavage stimulation factor (CstF), polyA polymerase, symplekin and pol II. The length of the polyA tail is controlled by PABP, which regulates the interaction between CPSF and PAP[88]. The cleavage occurs within a polyadenylation signal (PAS), a well characterized motif[87], which resides 10-30 nucleotides upstream to the cleavage site. CPSF binds the PAS directly. There is also a U/GU-rich region 30 nucleotides downstream to the cleavage site, which associates with CstF. The canonical PAS is AAUAAA or AUUAAA. There are additional sequences that can contribute to the efficiency of the 3' end processing, upstream and downstream to the cleavage sites. Upstream sequences are generally U-rich signals, which recruit 3' end processing factors and serve as another anchor for the 3' end processing machinery. Downstream sequences are G-rich, and they bind regulatory factors to enhance the process[88].

In recent years, it was discovered that most mRNAs contain multiple PASs with different affinities to the cleavage machinery, therefore creating alternative polyadenylation sites (APAs). The distal PAS (which creates the longest 3'UTR) tends to be the canonical PAS, with high affinity to the cleavage machinery, and is generally well conserved. Proximal PASs tend to diverge from the canonical motif and are less conserved[89]. Alterations in the 3' end processing pattern can lead to different diseases, such as oncological, immunological, neurological and hematological disorders[90].

Global mapping of APAs became possible in recent years due to advances in sequencing techniques. The study of APAs started from ESTs and microarrays, and developed further thanks to RNA-seq, and today thanks to 3′-end-enriched RNA-seq[87].

The usage of APA sites versus the canonical PASs is influenced by a variety of factors. The strength of the PAS is an important feature, but not the only one. PAS selection is also modulated by the levels of the 3' processing machinery. For example, during B-cell activation, CsF64 - a core 3' processing factor - is highly expressed, which leads to the usage of proximal and weak PAS in the *IgM* mRNAs[89]. RNA binding proteins also play a role in APA selection. They can mask directly the proximal PAS, as exemplified by PABPN1, which was shown to bind to APA sites and render them inaccessible to the cleavage machinery[91]. Other RNA binding proteins, such as HuR and ESRP1/2, were also shown to directly compete with the cleavage machinery on the alternative PASs. Another type of APA regulators is factors that bind outside the PAS and modulate their usage. One such example is the U1 snRNP, which binds the 5′ splice site and suppresses cleavage/polyadenylation at downstream PAS[89].

The usage of the canonical PAS is the default and is most common, while the proximal PASs play a role in regulation and are differentially used. The usage pattern of APA is highly conserved, and specific cell types and conditions appear to be more prone to APA usage than others[89]. During cell proliferation, there is a shift towards APA usage and global shortening of 3'UTRs[87]. Elkon et al. compared the 3'UTR end profiles of two cell lines during proliferation and arrest, and observed a broad induction of proximal PASs during the proliferative state, which they connected to E2F-mediated co-transcriptional regulation[92]. Cancerous cells appear to shorten the 3' UTR even more than non-cancerous proliferative cells[93]. Lin et al. generated polyadenylation maps of five tumor-normal pairs from different tissues, and observed a preferential usage of APA[94]. Lembo et al. found that tumors expressing shorter 3'UTRs tend to be more aggressive and have worse prognosis[95]. The transition from differentiated cells to induced pluripotent stem cells is also accompanied by global 3'UTR shortening[96]. A tendency to use proximal APA sites is seen also in particular tissues, such as placenta, ovaries and blood[97]. In contrast, processes such as embryonic development and myogenic differentiation of cultured myoblasts are accompanied by progressive lengthening of 3'UTRs[87].

The functionality of APA is the subject of many studies. The 3'UTR contains many cis elements, recognized by a variety of RNA binding proteins that regulate the biology of the mRNA, including subcellular localization, half-life and the rate of translation. These signals can be lost during APA. Asymmetry in localization of

mRNAs is important for cell polarity, division and more. For example, in the 3'UTR of BDNF (a neurotrophin important for synaptogenesis and synaptic plasticity), there is a localization element that leads the mRNA to dendrites. The short 3'UTR form of BDNF does not include this localization element, and consequently the mRNA is localized to the soma of neurons[98].

An interesting aspect of 3'UTR shortening is the interplay with miRNA binding sites. Mayr and Bartel found that in cancer cells, some APA isoforms of proto-oncogenes with shorter 3'UTRs tend to be more stable, generate more protein and lead to higher neoplastic transformation rates. This effect was due to their ability to escape miRNA regulation, as the corresponding binding sites were eliminated with the usage of the shorter 3'UTR form[93]. This notion of escaping miRNA regulation due to APA was shown also in other cases. *ABCG2* mRNA escapes regulation by miR-519c in its short form in drug-resistant cells[99]. *Hsp70* mRNA is alternatively polyadenylated upon ischemia or heat shock, and therefore escapes miR-378* regulation[100]. PAX3, a myogenic regulator, loses its miR-206 binding site when there is usage of an APA site in its 3'UTR in quiescent muscle stem cells[101].

The location of the binding site along the 3'UTR contributes to its functionality[3]. Conserved and functional binding sites tend to be preferentially positioned near the beginning and end of the 3'UTR[3]. In a paper recently published by the Bartel lab, they compared the impact of miRNAs in different cell lines. Their conclusion was that the binding sites repertoire was largely unaffected between cell lines, and the outliers that did change were due to APA. They determined that APA influenced 10% of the predicted binding sites when comparing any two cell lines. The major effect was due to binding site elimination by using a shorter 3'UTR isoform. However, they did mention that the distance of the binding site from the 3'UTR end can also vary due to APA. They compared the downregulation of genes with different 3'UTR lengths between two cell lines, where in both cell lines the miRNA binding site is included but at different distances from the 3'UTR end, and found that when the binding site was closer to the end the mRNA was more efficiently repressed by the miRNA[102].

3. The majority of endogenous microRNA binding sites within Alu elements avoid the microRNA machinery

Alus are repetitive elements, spread all over the human genome, and are present in many genes' '3UTRs. They have the ability to be inserted into new locations as well. They are about 300 bases long, and therefore have many potential miRNA binding sites within their sequence. In my research, I investigated whether Alus can spread miRNA regulation in the human genome. I found that most miRNA binding sites within Alus are non-functional and are avoided by the miRNA machinery. Moreover, I revealed 3 mechanisms that prevent such sites from being functional.

Bioinformatics. 2013 Apr 1;29(7):894-902. doi: 10.1093/bioinformatics/btt044. Epub 2013 Jan 29.

The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery.

Hoffman Y[1], Dahary D, Bublik DR, Oren M, Pilpel Y.

⊕ **Author information**

Abstract

MOTIVATION: The massive spread of repetitive elements in the human genome presents a substantial challenge to the organism, as such elements may accidentally contain seemingly functional motifs. A striking example is offered by the roughly one million copies of Alu repeats in the genome, of which ~0.5% reside within genes' untranslated regions (UTRs), presenting ~30 000 novel potential targets for highly conserved microRNAs (miRNAs). Here, we examine the functionality of miRNA targets within Alu elements in 3'UTRs in the human genome.

RESULTS: Using a comprehensive dataset of miRNA overexpression assays, we show that mRNAs with miRNA targets within Alus are significantly less responsive to the miRNA effects compared with mRNAs that have the same targets outside Alus. Using Ago2-binding mRNA profiling, we confirm that the miRNA machinery avoids miRNA targets within Alus, as opposed to the highly efficient binding of targets outside Alus. We propose three features that prevent potential miRNA sites within Alus from being recognized by the miRNA machinery: (i) Alu repeats that contain miRNA targets and genuine functional miRNA targets appear to reside in distinct mutually exclusive territories within 3'UTRs; (ii) Alus have tight secondary structure that may limit access to the miRNA machinery; and (iii) A-to-I editing of Alu-derived mRNA sequences may divert miRNA targets. The combination of these features is proposed to allow toleration of Alu insertions into mRNAs. Nonetheless, a subset of miRNA targets within Alus appears not to possess any of the aforementioned features, and thus may represent cases where Alu insertion in the genome has introduced novel functional miRNA targets.

SUPPLEMENTARY INFORMATION: Supplementary data are available at Bioinformatics online.

PMID: 23361327 DOI: 10.1093/bioinformatics/btt044

3.1 Materials and Methods

Human and mouse genomes and repeat sequences: The full human 3'UTR sequence dataset was taken from UCSC[103] NCBI36/ hg18. Alu sequences and their locations were taken from Repeat Masker (http://www.repeatmasker.org). The mouse 3'UTR sequences and coordinates were taken from UCSC[103] (NCBI37/mm9). The mouse repeats were taken from UCSC Repeat Masker (http://www.repeatmasker.org).

Prediction of miRNA binding sites and conservation analysis: miRNA binding sites were predicted by scanning for the seed of the miRNA, on the basis of perfect (Watson–Crick) complementarity. Binding sites were defined as perfect 7-mers, for all human and mouse miRNAs listed in miRBase release 15 (http://www.mirbase.org/)[104]. Conserved binding sites were taken from TargetScan release 5.1[3], 7mer-m8 binding site type only (perfect 7-mer). The analysis of conservation and folding energy included only genes with the following attributes: (i) the 3'UTR in the UCSC version was the same as the 3'UTR used by TargetScan (as defined in their website) and (ii) the 3'UTR from UCSC was included within the 3'UTR defined in TargetScan or the opposite.

Analysis of miRNA overexpression data: The data of miRNA overexpression experiments were taken from Khan et al. 2009[105], a database that contains siRNAs as well as miRNA overexpression experiments. A subset of 43 miRNA overexpression experiments was analyzed. For each overexpressed miRNA, its site was scanned against all human 3'UTRs. Downregulation of binding site mRNAs was defined as the percentage of genes with fold reduction of at least 1.62 (i.e. 0.7 on a log2 scale). The cut-off was decided according to the distributions of average fold change for genes with and without the miRNA binding site. For each analysis, only experiments where the group of genes for consideration consisted of at least eight genes were included.

Secondary structure prediction: Secondary structures were predicted for all human and mouse 3'UTRs using the Bioinformatics Toolbox of Matlab 10, which implements the M-Fold and Vienna algorithms[106,107]. The analysis was done in windows of 100 bp, and up to 50 bps from the last coding exon were added to the beginning of the 3'UTR for the prediction. The secondary structure status 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.

Analysis of PAR-CLIP data: The raw data of the PAR-CLIP experiment were taken from Kishore et al., 2011[108]. The representation of the most abundant miRNAs in the PAR-CLIP data is highly correlated between replicates and between RNAse protocols. The experiment with the highest number of mapped reads (GSM714644) was further analyzed. The reads were mapped to the genome using Bowtie[109] (with the parameter 'best' so that for each read, we received only one mapping, and the hg18 genome index). The Bowtie output was filtered to include only reads with five mutations or less. The sequence of the reads was corrected to the genome. The miRNA binding sites were identified in the reads according to the reads' genome-corrected sequence. The decision of which read is within Alu was according to the location of its best mapping. We did not attempt to map the Alu reads to their exact location in the genome (as the best mapping of Alu reads are probably one of many best mappings, as they appear in many locations in the genome), but simply infer from their best mapping if they are Alus or not.

Calculating miRNA binding site representation in the transcriptome: For each miRNA of the 10 most abundant miRNAs in the PAR-CLIP experiment, the percentage of expressed miRNA binding sites within Alus was calculated, according to the mRNA-Seq experiment done by Kishore et al., 2011[108] (GSM714678 and GSM714679, which mimic best the conditions of the PAR-CLIP experiment). In each transcript, the numbers of miRNA binding sites, in total and within Alus, were calculated, and multiplied by the count of the transcript. Transcripts with low count (<10) were excluded. The average of the two replicates is represented in the analysis.

3.2 Results

3.2.1 Potential genomic interplay between Alus and microRNAs

Alus present a substantial collection of miRNA binding sites within genes' 3'UTRs. 16% of human genes contain at least one Alu element within their 3'UTR. A total of 4,927 Alu sequences that reside within 3'UTRs present 94,785 potential miRNA binding sites (defined as 7-mers with perfect match to positions 2–8 of the miRNA), 28,829 of which correspond to a set of 401 miRNAs that are conserved among mammals[110]. To examine the effects of Alu insertions in 3'UTRs on gene expression, I analyzed data from reported miRNA overexpression experiments. Khan et al. have assembled the results of dozens of such genome-wide expression array and proteomics experiments into a single normalized database[105]. For subsequent analysis, I used a subset of experiments from the Khan database, comprising 43 experiments with 23 different miRNAs overexpressed in a total of five different cell lines. For each experiment, the dataset provides the genome-wide mRNA response to the overexpression of one miRNA at a time in a given cell line. As a preliminary step, we assessed the potential of this dataset to demonstrate known attributes of miRNA regulation. First, we examined whether genes that contain a predicted binding site for a particular miRNA are more likely than other genes to be downregulated in response to overexpression of that miRNA. Reassuringly, Figure 1A shows that the percentage of downregulated genes was significantly higher in the group of genes that contain a putative binding site for the overexpressed miRNA, relative to the group of genes lacking such binding site (p-value=2.9e-21, Student's t-test). Moreover, as already suggested by others[3,111], genes with more than one putative binding site for a given miRNA are more efficiently downregulated than genes with only a single binding site (p-value=5.5e-5, Student's t-test; Figure 1B). The mere existence of a miRNA binding site sequence inside the 3'UTR of a gene does not necessarily imply that the gene will constitute a functional target for the miRNA. A commonly accepted hallmark of a binding site's authentic functionality is its evolutionary conservation[3,112,113]. Therefore, we next compared between genes that contain a conserved versus non-conserved binding site for each overexpressed miRNA. Figure 1C shows that the group of genes with conserved binding sites has a significantly

larger percentage of downregulated genes, relative to human genes that contain the same binding site, but this motif is not conserved in the orthologous genes of other mammals (p-value=5.6e-14, Student's t-test). Together, these results demonstrate that the miRNA overexpression dataset reflects faithfully established parameters of miRNA regulation and can therefore be further used to examine other features such as the effect of miRNA motifs inside Alu sequences.

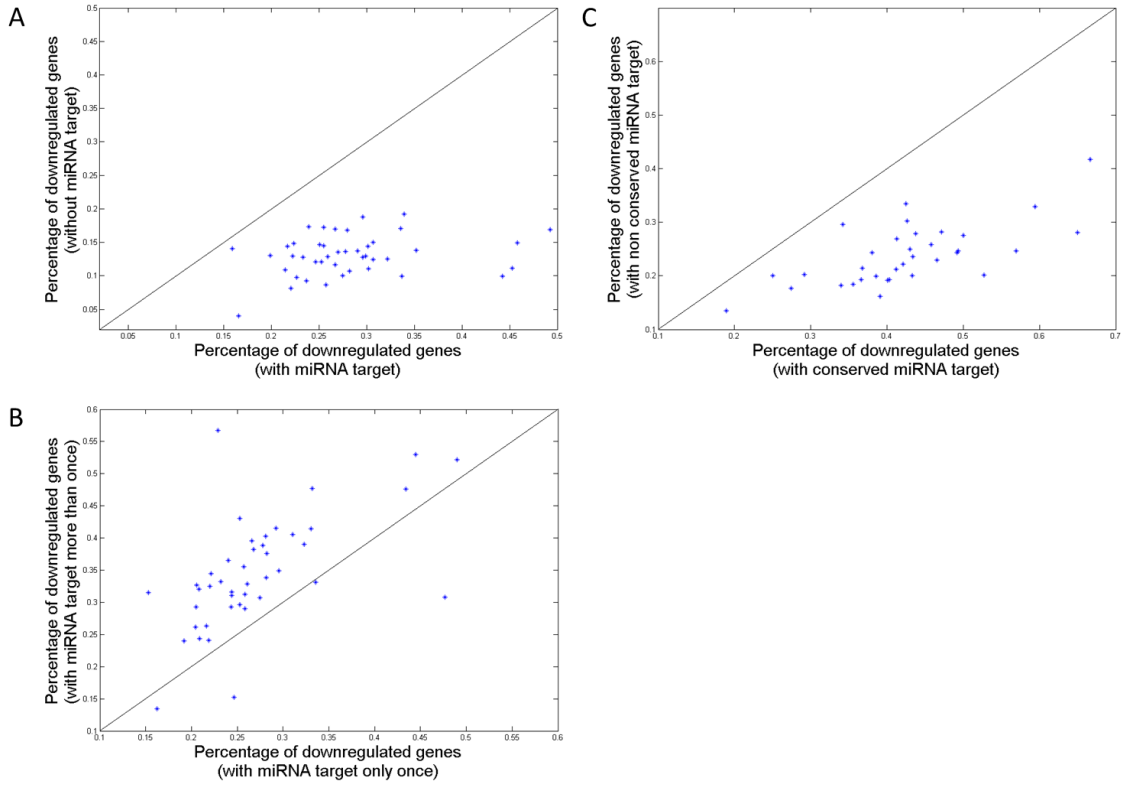


Figure 1: Features of miRNA binding sites in the Khan dataset

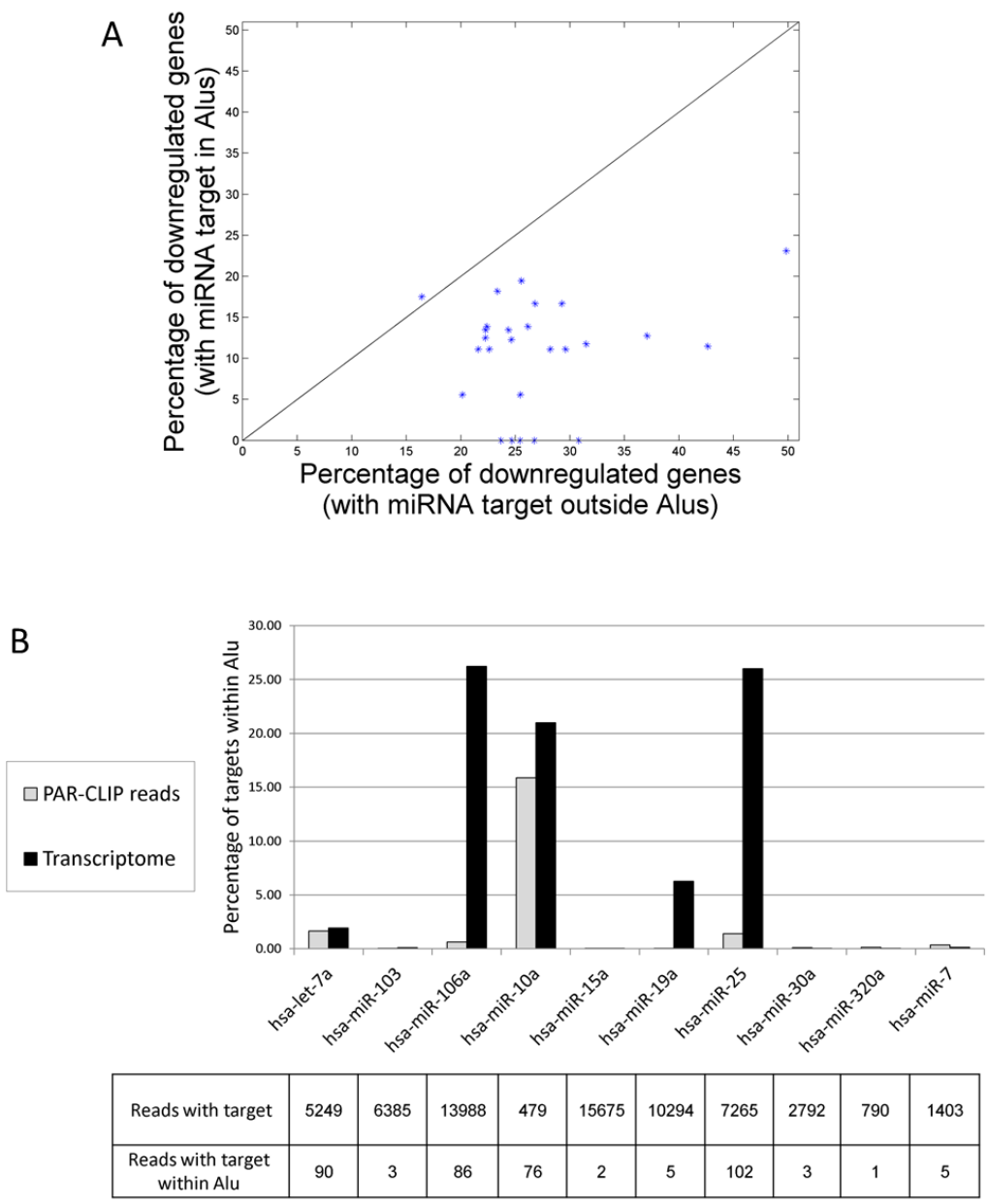
Each dot represents a single experiment in which a certain miRNA was overexpressed. Each experiment is plotted according to the percentage of downregulated genes in each group: genes with miRNA binding sites vs. without (A), Genes with a single miRNA binding site vs. multiple binding sites (B), and genes with a conserved vs. non conserved binding site (C).

3.2.2 miRNA binding sites inside Alus are less functional

To discriminate between the effects of miRNA binding sites inside versus outside Alus, we defined for each miRNA in the dataset two sets of genes: the first consists of genes in which the putative miRNA binding site resides only inside Alus and the second contains the site only outside Alus. We found that genes with miRNA binding sites only outside Alus are more likely to be downregulated than genes with binding sites only inside Alus (Figure 2A, p-value=2.2e-11, Student T-test).

The compromised functionality of putative miRNA binding sites within Alus can be due to a failure of the miRNA machinery to bind such sites, or to its dysfunctionality after the binding has occurred. To distinguish between these two possibilities, I analyzed data from Ago2-mRNA binding experiments[108]. Ago2 is part of the Argonaute family of proteins, which are guided by the mature miRNA to bind the specific complementary region of the target mRNA in order to initiate its silencing[114]. Therefore, profiling of Ago2-bound mRNA species could serve as a means to assess how efficiently a given RNA sequence is bound by the RISC machinery. Analysis of PAR-CLIP data supports the notion that miRNA binding sites within Alus are not functional, since they are not even bound to the Ago machinery. Notably, miR-106a, which is highly expressed in the cells, has 27% of its expressed potential binding sites within Alus, but strikingly only 0.61% of its associated reads could actually be mapped to Alus (Figure 2B, p-value<e-300, HyperGeometric test).

These results suggest that miRNA binding sites within Alus are not functional because they fail to even bind the miRNA machinery, and that insertions of Alu-contained miRNA sites into 3'UTRs were largely tolerated only when they could escape Ago2 binding and hence silencing.



Reads with target	5249	6385	13988	479	15675	10294	7265	2792	790	1403
Reads with target within Alu	90	3	86	76	2	5	102	3	1	5

Figure 2: miRNA binding sites within Alus are less functional

(A) Each dot represents a single experiment in which a certain miRNA was overexpressed. Each experiment is plotted according to the percentage of downregulated genes in each group: genes with miRNA binding sites only outside Alus versus only within Alus. (B) For each of the 10 miRNAs identified by the PAR-CLIP experiment as being most abundant in the analyzed cells, the percentage of predicted binding sites within Alus is compared between the overall transcriptome and the PAR-CLIP reads. Bellow the x axis, the table depicts the absolute number of PAR-CLIP reads containing the putative miRNA binding site.

3.2.3 Three mechanisms to prevent the functionality of miRNA binding sites within Alus

Realizing that legitimate 7-mer perfect match binding sites within Alus are often not functional, I looked for features that would explain such lack of functionality. Conversely, absence of such features in exceptional cases may highlight potential novel functional miRNA binding sites that were inserted through Alu retro-transposition.

The first feature I explored is the location of Alu insertions within 3'UTRs. It was previously shown that conserved and functional miRNA binding sites tend to reside at both ends of 3'UTRs, and less in the UTR's middle[3]. I recapitulated this finding for relative position along the 3'UTR, showing that conserved miRNA binding sites are enriched near the two ends of the 3'UTR. Examining only 3'UTRs longer than 1000 bps, I found that conserved miRNA binding motifs are concentrated in the first and last 250 bps of the UTR, and are relatively depleted from the middle section. I next examined the positions of Alu insertions into 3'UTRs. Strikingly, I found that the majority of Alu insertions have occurred in the middle sections of the 3'UTR, presenting an almost exact mirror image of the conserved miRNA recognition motifs' location (Figures 3A,B). These results suggest that Alu insertions into 3'UTRs during primate evolution were tolerated, provided that they occurred away from the two 3'UTR ends.

Another feature that I found as a possible explanation for exclusion of Alu-residing miRNA target sites is the structure of the binding site. The RNA structure and folding energy of miRNA binding sites and their surroundings are important for their functionality; in particular, binding sites located within mRNA regions possessing tight secondary structure are typically less functional[5,6]. Indeed, I found that binding sites inside Alus tend to reside within tighter structures, as compared with conserved miRNA binding sites (Figure 3C).

Alu sequences are subject to extensive RNA editing[115-119], which modifies adenosines (A) to inosines (I). In fact, the majority of editing events in human tissues occur within Alus[120]. Since inosines are recognized as guanosines by many of the molecular machineries in the cell, such alterations can diminish the complementarity between a miRNA's seed and its binding site within Alus, or introduce novel binding sites by creating complementarity with the miRNA's seed[121]. When examining only

genes with putative miRNA binding sites without adenosine, I found that binding sites within Alus are moderately less functional than binding sites outside Alu (Figures 3D, p-value=0.02, Student's t-test), probably due to the effects of territory and secondary structure discussed above. Notably, when I examined only binding sites with adenosine, the binding sites within Alus were found to be substantially less functional than those outside Alus (Figures 3E, p-value=5.8e-6, Student's t-test). This observation suggests that among Alu-contained sites, miRNA binding sites without adenosine are more effective than adenosine-containing ones.

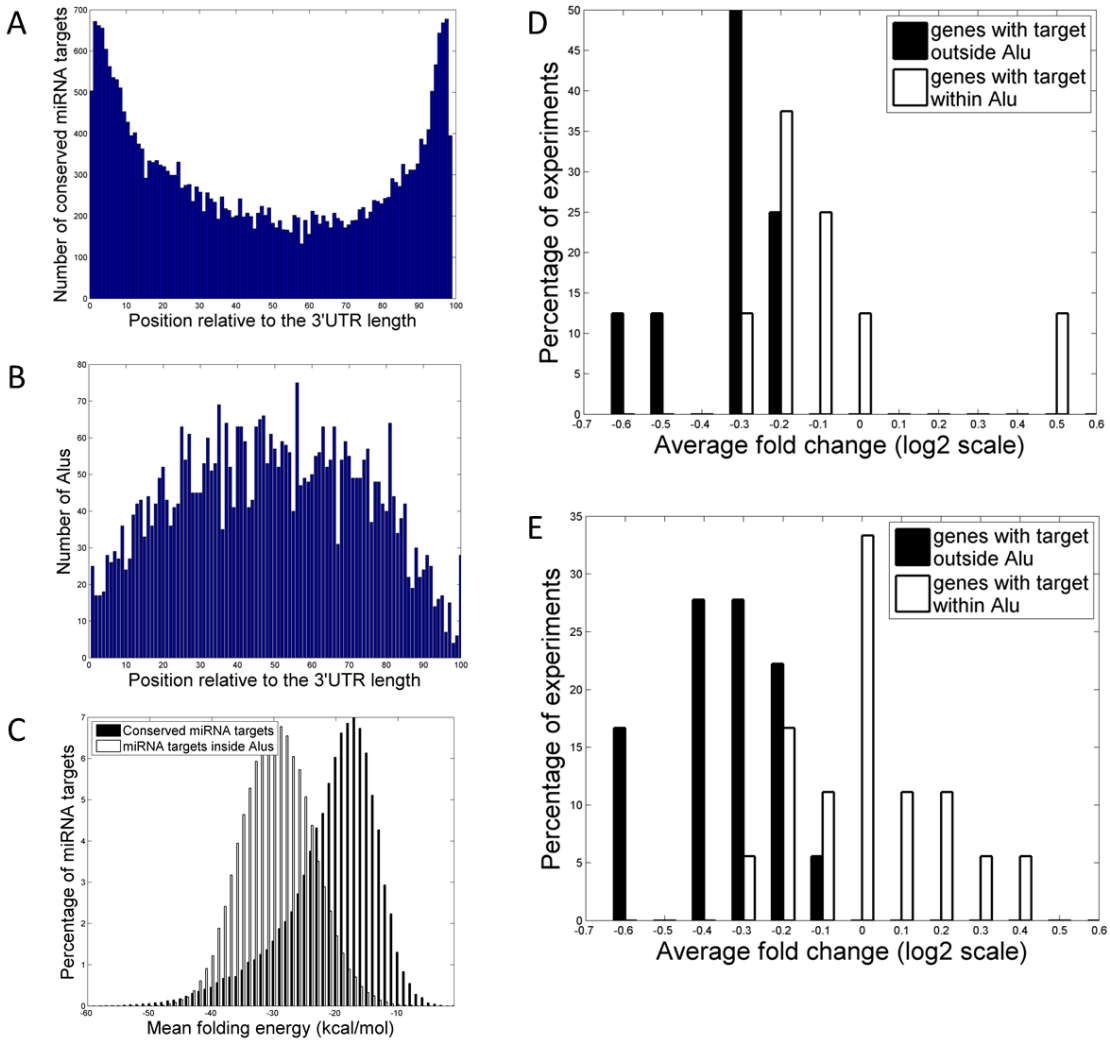


Figure 3: Three mechanisms to prevent miRNA binding sites within Alus from functionality

(A,B) Alus and miRNA binding sites territories within 3'UTRs: (A) distribution of conserved miRNA sites along 3'UTRs and (B) distribution of Alus along 3'UTRs. Only 3'UTRs longer than 1Kb were analyzed. The X-axis depicts the relative position on the 3'UTR (normalized to its length). (C) Distribution of the folding energy around conserved miRNA binding site sites and miRNA sites within Alus. (D,E) Average fold change of genes that contain the miRNA binding site within and outside

Alus, for over-expression experiments with miRNAs that do not contain an A in their binding site motif (D), or contain an A (E).

3.2.4 Repeats and miRNA binding sites in the mouse genome

Transposable elements are active in most animal genomes. Therefore, introduction of novel miRNA binding sites via transposition can occur in other species as well. Although Alu repeats are primate specific, the mouse genome also contains repeats similar to Alus, namely B1 repeats that belong to the Short Interspersed Elements family. As in the case of Alus, B1 repeats emerged from the ancestral 7SL RNA gene[122]. The B1 repeats are less widespread than Alus, comprising only 2.7% of the mouse genome[17,123] and are also shorter (~140 bp)[41]. In the mouse genome, 8.3% of the genes contain at least one B1 repeat in their 3'UTR. The 1,962 B1 sequences that reside within 3'UTRs represent 14,372 potential miRNA binding sites (perfect 7-mers). Consequently, the potential effect of putative miRNA binding sites within B1 repeats is less substantial than that of Alus in the human genome; it is nonetheless not negligible. I observed that with regard to their location within the 3'UTR, mouse B1 repeats show similar trends as the human Alus. B1 repeats tend to avoid the two ends of the 3'UTR, predominantly the beginning, while the conserved mouse miRNA binding sites display an opposite trend of preferential location near the UTR ends (Figure 4A). Furthermore, putative miRNA binding sites within B1 repeat-encoded mRNAs show tighter local secondary structure (Figure 4B). The fact that this feature is shared with Alus is probably explained by the common evolutionary origin of these two types of repeats[122]. In conclusion, like Alus in the human genome, B1 insertions into mouse mRNAs were probably tolerated, provided that they occur into 3'UTR territories that do not overlap with functional miRNA binding sites or that they possess tight secondary structure.

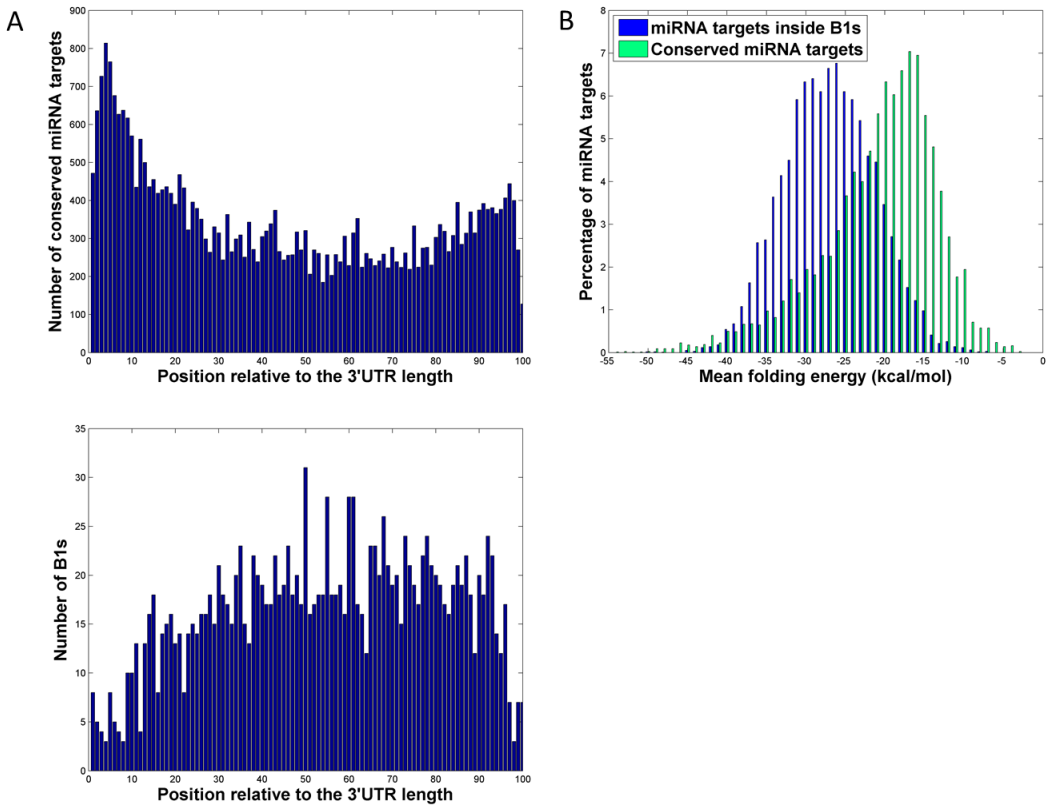


Figure 4: Repeats and miRNA binding sites in the mouse genome.

(A) The histograms show the number of conserved miRNA binding sites and number of B1s, at various relative distances from the beginning of the 3'UTR. The x-axis shows a scaling of the 3'UTR length so that UTRs of various lengths can be compared by normalizing for each UTR the location of a miRNA or Alu to the length of the corresponding UTR. Only 3'UTR with length above 1000 bases were considered. (B) The mean folding energy of all the miRNA binding sites inside B1 sequences and conserved binding sites in the mouse genome was calculated using Matlab.

4. miR-661 downregulates both Mdm2 and Mdm4 to activate p53

p53 is a highly important protein in cancer prevention. It has two major regulators – Mdm2 and Mdm4, which downregulate its expression and functionality. In my research I found a human miRNA – miR-661, which downregulates both Mdm2 and Mdm4, and upregulates p53 functionality. Most of the binding sites for miR-661 within *Mdm2* and *Mdm4* 3'UTRs are within Alus, therefore it presents an opposite example to my previous conclusion, showing a rare case where primate-specific regulation within Alus can be functional and important, even within a conserved pathway such as p53.

Cell Death Differ. 2014 Feb;21(2):302-9. doi: 10.1038/cdd.2013.146. Epub 2013 Oct 18.

miR-661 downregulates both Mdm2 and Mdm4 to activate p53.

Hoffman Y[1], Bublik DR[2], Pilpel Y[3], Oren M[2].

Author information

[1]1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel [2] Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.
[2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.
[3]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.

Abstract

The p53 pathway is pivotal in tumor suppression. Cellular p53 activity is subject to tight regulation, in which the two related proteins Mdm2 and Mdm4 have major roles. The delicate interplay between the levels of Mdm2, Mdm4 and p53 is crucial for maintaining proper cellular homeostasis. microRNAs (miRNAs) are short non-coding RNAs that downregulate the level and translatability of specific target mRNAs. We report that miR-661, a primate-specific miRNA, can target both Mdm2 and Mdm4 mRNA in a cell type-dependent manner. miR-661 interacts with Mdm2 and Mdm4 RNA within living cells. The inhibitory effect of miR-661 is more prevalent on Mdm2 than on Mdm4. Interestingly, the predicted miR-661 targets in both mRNAs reside mainly within Alu elements, suggesting a primate-specific mechanism for regulatory diversification during evolution. Downregulation of Mdm2 and Mdm4 by miR-661 augments p53 activity and inhibits cell cycle progression in p53-proficient cells. Correspondingly, low miR-661 expression correlates with bad outcome in breast cancers that typically express wild-type p53. In contrast, the miR-661 locus tends to be amplified in tumors harboring p53 mutations, and miR-661 promotes migration of cells derived from such tumors. Thus, miR-661 may either suppress or promote cancer aggressiveness, depending on p53 status.

PMID: 24141721 PMCID: PMC3890952 DOI: 10.1038/cdd.2013.146

4.1 Materials and Methods

Cell culture, siRNA and miRNA transfections: Cells were maintained at 37^{o}C in DMEM (Biological Industries) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Hyclone) (besides OVCAR-8 cells which were maintained in RPMI with 5% heat-inactivated fetal bovine serum) and penicillin-streptomycin antibiotics solution (Biological Industries). Transient transfection of miRNA and siRNA was performed with Dharmafect 4 (MDA-MB-231) or Dharmafect 1 (all other cell lines) according to the manufacturer's (Dharmacon) instructions. miRNA mimics (Dharmacon) were used at a final concentration of 20 nM; siRNA (Dharmacon) was used at different concentrations. For RNA and protein analysis Dharmafect Smart-pool siRNA was used. For cell cycle analysis single siRNA oligos (Sigma and Dharmacon) were employed. miRNA inhibitor (miRIDIAN microRNA Hairpin Inhibitor; Dharmacon) was used at 100 nM final concentration.

RNA purification and Real-Time quantitative PCR: RNA was extracted with the mirVana miRNA Isolation Kit (Ambion). For quantitative reverse transcriptase-PCR (qRT-PCR) analysis, 0.7-1.5 μg of each RNA sample was reverse transcribed with Moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI), random hexamer primers (Sigma) and dNTPs (LAROVA). qRT-PCR was done in a StepOne real-time PCR machine (Applied Biosystems, Foster City, CA) with Syber Green PCR supermix (Invitrogen).

Primers: The following primers were used (Sigma):

Gene	Forward primer	Reverse primer
p53	CCCAAGCAATGGATGATTTGA	GGCATTCTGGGAGCTTCATCT
GAPDH	AGCCTCAAGATCATCAGCAATG	CACGATACCAAAGTTGTCATGGAT
MDM4	AATGATGACCTGGAGGACTCTA	ACTGCCACTCATCCTCAGAGGTA
p21	GGCAGACCAGCATGACAGATT	GCGGATTAGGGCTTCCTCTT
MDM2	CAGGCAAATGTGCAATACCAA	GGTTACAGCACCATCAGTAGGTACAG
CD95	CCCTCCTACCTCTGGTTCTTACG	TTGATGTCAGTCACTTGGGCAT
Btg2	CCAGGAGGCACTCACAGAGC	GCCCTTGGACGGCTTTTC
Wig1	AGCTGTCCTCCTCCTGCTAGAA	TCTGCGGAGGGACTGGAAC
Actin	CATGAAGATCAAGATCATCGCC	ACATCTGCTGGAAGGTGGACA

Antibodies: The following primary antibodies were used for Western blot analysis.

GAPDH: monoclonal antibody Millipore MAB374; Mdm2: monoclonal antibodies 4B2, 2A9, and 4B11; Mdm4: BL1258 (Bethyl Laboratories); p53: monoclonal antibodies PAb18O1 and DO1; p21:c-19 (Santa Cruz Biotechnology).

Western blot analysis: For Western blot analysis cells were washed with PBS, collected and lysed with NP40 lysis buffer (150 mM Sodium Chloride, 50 mM Tris pH=8, 1% NP40) with protease inhibitor cocktail (Sigma). Cells were vigorously vortexed and centrifuged at 14,000 RPM for 10 minutes at 4°C, and the soluble fraction was used to determine protein concentration in each sample. The protein concentration was quantified with the BCA kit (Thermo scientific) according to the manufacturer's protocol. Protein sample buffer (3% SDS, 10% glycerol, 5% β-mercaptoethanol, 62 mM 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 PBS-T (0.05% Tween-20 in PBS), and reacted for 45 minutes with horseradish peroxidase (HRP)- conjugated IgG, followed by 3 washes with PBS-T and one wash in PBS. The proteins were visualized using an enhanced chemiluminescence (ECL) detection kit (Amersham), followed by exposure to X-ray film or analysis in a ChemiDoc MP imaging system (Bio RAD). Bands were quantified with Image Lab 4.1.

Luciferase assays: Cells were seeded in 12-well dishes and transfected with miRNA (20 nM final) as described above. 48 hours later, cultures were transfected with 200 ng of firefly luciferase reporter plasmid DNA (p21 WT or p21 mutated) and 40 ng renilla luciferase plasmid DNA, using the JetPEI reagent (Polyplus Transfection) in NaCl, according to the manufacturer's protocol. 24 hours later 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).

BrdU incorporation analysis: 24 hours after seeding, cells were transfected with miR-661 or miR-control (20 nM). 24 hours later, cells were retransfected with 20 nM p53 siRNA (Sigma or Dharmacon single oligos) or LacZ siRNA (Dharmacon). After additional 24 hours, cells were analyzed for BrdU incorporation as previously described[124,125].

Colony formation assays: 24 hours after seeding, cultures were transfected with 20 nM miR-661 or miR-control. After additional 24 hours, cells were counted, seeded in a 6-well plate at a density of 3,000 cells/well and incubated for 8 days at 37°C in a humidified atmosphere of 5% CO2. The colonies were fixed with cold methanol for 5 min, stained with 0.1% crystal violet for 10 min and washed with distilled water.

miRNA pull-down: miRNA pull-down assays were performed as described [126,127]. MCF7 cells were seeded in 10 cm dishes 24h before being transfected with biotinylated miR-661 mimic or miR control (100 nM; Dharmacon). After 48h, cells were harvested in lysis buffer [20 mM Tris pH 7.5, 100 mM KCL, 5 mM MgCl2 and 0.3% NP-40, including 100 u/ml RNAse inhibitor (Promega) and Protease Inhibitor mix (Sigma)] and incubated with Streptavidin Dynabeads (Invitrogen) for 4 hours at 4°C with constant rotation. The beads were prepared and washed according to the manufacturer's instructions, and incubated for 1 hour at 4°C with lysis buffer including 1 mg/ml RNAse-free BSA and 1 mg/ml yeast tRNA (both from Ambion) prior to incubation with the lysed cells. After incubation with the beads, 2 washes with lysis buffer were performed and RNA was extracted with Trizol (Invitrogen) and Chloroform (Fisher Scientific). cDNA preparation and qRT-PCR were done as described above, and values were normalized to input (cellular RNA without incubation with beads) and then to *GAPDH*.

Cell migration analysis: Cell migration was evaluated with the aid of a real-time cell analyzer (xCELLigence RTCA; Roche), which provides a real-time measurement of migrating cells by extrapolating changes in electrical impedance with the number of cells passing through a porous membrane. Briefly, 160 μL of complete RPMI medium supplemented with 10% FBS (as attractant) were loaded in the lower chamber of the migration plate (CIM-Plate 16; Roche). After fitting the upper chamber on the lower chamber, 35 μL of RPMI containing 0.1% FBS were loaded and allowed to equilibrate for 1 hour in a 37°C incubator. 8×10^4 OVCAR-8 cells, transfected 48 h earlier with miR-C or miR-661 (20 nM) or with si-miR-C or si-miR-661 (100 nM), were starved for 6 h in RPMI without FBS and then resuspended in 100 μl of RPMI containing 0.1% FBS. Then, cells were loaded in the wells of the upper chamber in the CIM-Plate (subsequently placed in the RTCA analyzer in a 37°C incubator). After background reading was determined, cell migration was measured and recorded every 15 minutes (100 sweeps at 15-minute intervals). RPMI medium without FBS loaded

in the lower chamber was used as negative control. Each experiment was performed in three biological replicates.

Clinical data analysis: Data of miR-661 amplification and p53 status in patients from different cancers was generated using cBio portal[128]. For each cancer type the percentage of patients with miR-661 amplification, p53 mutation or both together was calculated as an average of all datasets available for this cancer type. P-Values were calculated with a hypergeometric distribution. Survival probabilities of breast cancer patients were generated with MIRUMIR[129], based on data taken from GSE37405[130].

4.2 Results

4.2.1 miR-661 downregulates simultaneously both Mdm2 and Mdm4

miR-661, a primate-specific miRNA, is not well studied and is known to have only a few validated binding sites, all outside the p53 network. *Mdm2*'s 3′UTR contains three potential binding sites for miR-661 all within Alus, and *Mdm4* contains nine binding sites, all but one within Alus (Figure 5).

To investigate whether miR-661 can indeed target *Mdm2* mRNA, I transiently transfected MCF7 breast cancer cells as well as other cell lines with miR-661 mimic. Although this led to only a slight reduction in *Mdm2* mRNA (Figure 6A), Mdm2 protein levels were markedly downregulated (Figure 6B). The effect of miR-661 on Mdm2 protein levels was reproduced in a variety of other cell lines, including A549 and H460 (non-small cell lung cancer) (Figure 6B). In most cases, p53 protein levels were not altered. Interestingly, while *Mdm2* mRNA levels were only marginally affected in MCF7 and A549 cells, a more significant reduction could be observed in MDA-MB-435 breast cancer cells (Fig. 6A). Of note, MCF7 and A549 express wild type p53 (WTp53) whereas MDA-MB-435 express mutant p53. Since Mdm2 downregulation is expected to increase the transcriptional activity of p53, and the *Mdm2* gene is a positive transcriptional target of p53, the negative effect of miR-661 on *Mdm2* mRNA may be partly compromised, in cells expressing WTp53, by increased transcription of the *Mdm2* gene. To address this possibility, I doubly transfected MCF7 cells with a combination of miR-661 mimic and p53 siRNA (sip53), thereby attenuating the Mdm2-p53 feedback loop. Indeed, depletion of p53 revealed a stronger downregulation of *Mdm2* mRNA by miR-661 (Figure 6C); in agreement, the decrease in Mdm2 protein was also more pronounced (Figure 6C). Finally, to validate that endogenous miR-661 also targets Mdm2, I transfected MCF7 cells with miR-661 inhibitor. As seen in Figure 6D this led to an increase, albeit modest, in Mdm2. In sum, these observations identify Mdm2 as a *bona fide* target of miR-661.

As noted above, miR-661 is predicted to target also Mdm4. Indeed, transfection of MCF7 cells with miR-661 mimic elicited a modest reduction in Mdm4 protein (Figure 6E), although we did not observe a significant effect on *Mdm4* mRNA (data

not shown). Unlike its widespread effect on Mdm2, miR-661 did not suppress Mdm4 protein in several other cell lines (data not shown), suggesting that its ability to target *Mdm4* mRNA is highly context-dependent. Conceivably, miR-661 may regulate Mdm4 in synergy with other miRNAs, expressed in MCF7 but not in the other cell lines examined. In fact, the *Mdm4* 3'UTR is exceptionally long, and is predicted to harbor binding sites for a multitude of miRNAs.

Of the 9 predicted miR-661 targets within the *Mdm4* mRNA 3'UTR, all except one reside within Alu repeats (Figure 5). Since Alu-embedded miRNA binding sites are often non-functional[131], we surmised that the single non-Alu binding site was responsible for inhibition by miR-661. However, when cloned in a luciferase reporter, a 300 base pair fragment spanning this binding site had no detectable effect in MCF7 cells (data not shown), as was also the case when several Alu-embedded putative binding sites were similarly tested individually. Thus, a combination of two or more binding sites may be required to mediate the inhibitory effect of miR-661 on Mdm4.

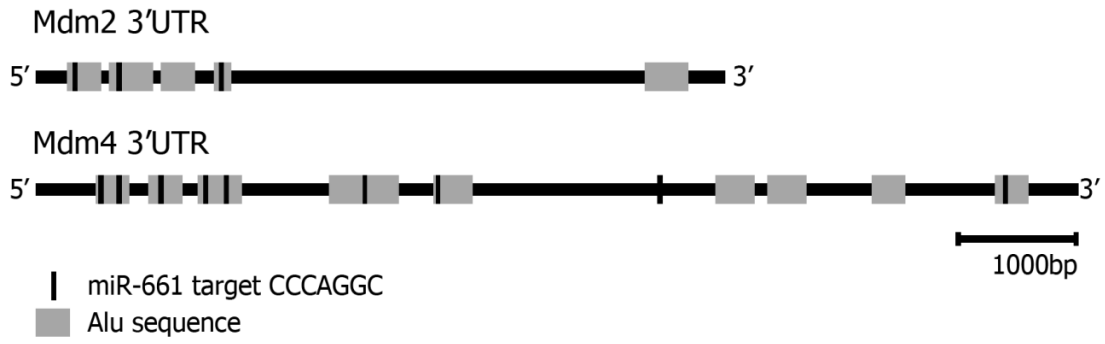


Figure 5: *Mdm2* and *Mdm4* mRNA contain multiple predicted targets for miR-661 in their 3'UTR

Schematic display of *Mdm2* and *Mdm4* 3'UTRs, and the positions of putative miR-661 targets and Alu sequences.

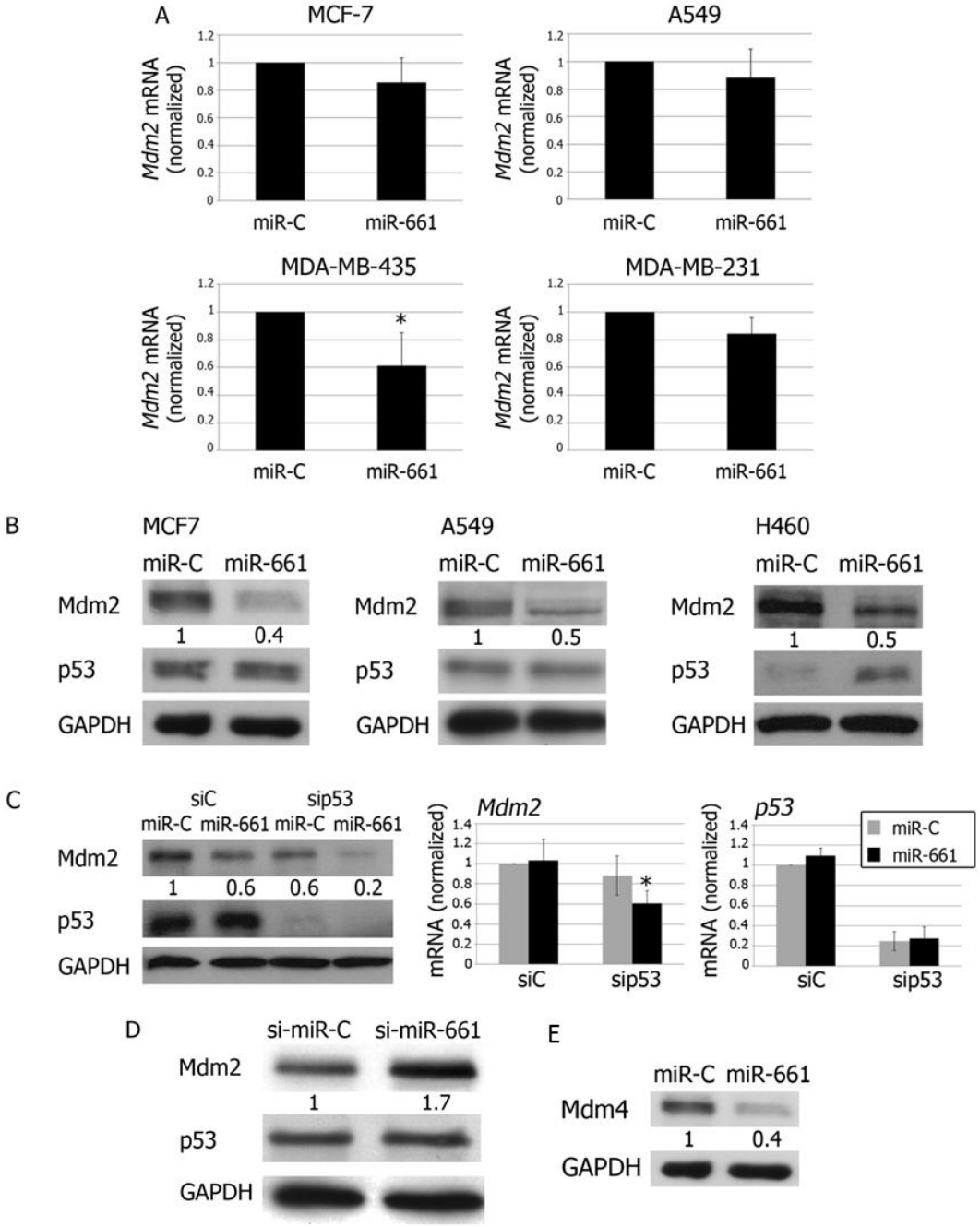


Figure 6: miR-661 downregulates Mdm2 and Mdm4

(A) The indicated cell lines were transfected with miR-661 mimic (miR-661) or miR-control (miR-C; 20 nM final), and harvested 48 hours later for RNA extraction and qRT-PCR analysis of Mdm2 mRNA. Values were first normalized to GAPDH mRNA in the same sample, and then calculated relative to the miR-C value, set as 1. Values represent the average +/- SD from 3 (MCF7), 2 (A549), 5 (MDA-MB435) and 4 (MDA-MB-231) independent experiments. P-Values for the difference between miR-661 and miR-C: MCF7=0.27, A549=0.6, MDA-MB-435=0.03, MDA-MB-231=0.1; Student's t-test) (B) Cells transfected as in (A) were lysed and subjected to Western blot analysis with the indicated antibodies. Mdm2 band intensities were quantified. Values were first normalized to GAPDH intensities in the same sample, then calculated relative to the miR-C value, set as 1. (C) MCF7 cells

were transfected with miR-661 mimic or miR-control (20 nM final) in combination with p53 siRNA (sip53; 20 nM) or control siRNA (siC; 20 nM). 48 hours later, cells were harvested for Western blot analysis with the indicated antibodies (left) or for RNA extraction and qRT-PCR analysis (middle and right panels) as in (A,B). P-Value for sip53 in the middle panel=0.04; Student's t-test) (D) MCF7 cells were transfected with miR-661 inhibitor (si-miR-661; 100 nM) or miR-control inhibitor (si-miR-C), and harvested 48 hours later for Western blot analysis with the indicated antibodies. (E) Cells were treated and harvested for Western blot analysis with the indicated antibodies as in (B). Mdm4 band intensity was quantified and calculated relative to GAPDH in the same sample and to the miR-C value, set as 1.

4.2.2 miR-661 interacts with *Mdm2* and *Mdm4* mRNA within cells

To obtain more direct evidence for the interaction of miR-661 with *Mdm2* and *Mdm4* mRNA, I performed a miRNA pull-down assay. Briefly, cells were transfected with biotinylated miR-661 mimic or miRNA control; biotinylated miR-661 retained the ability to downregulate both Mdm2 and Mdm4 (Figure 7B). Cell extracts were then prepared and reacted with streptavidin-coupled beads in order to purify the miRNA mimic together with its associated mRNA molecules[126,127]. As seen in Figure 7A, both *Mdm2* and *Mdm4* mRNAs were significantly enriched in the miR-661 pull-down relative to the miR-control pull-down; *Mdm2* mRNA displayed a greater fold enrichment than *Mdm4* mRNA. Actin mRNA, which is not a predicted miR-661 target, did not undergo comparable enrichment. These data strongly suggest that, as predicted computationally, miR-661 binds directly *Mdm2* and *Mdm4* mRNA.

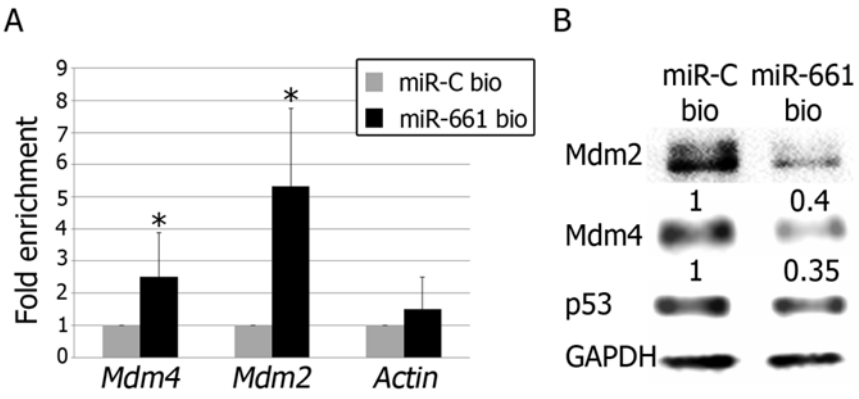


Figure 7: miR-661 binds Mdm2 and Mdm4 mRNA within cells

(A) MCF7 cells were transfected with biotinylated miR-661 mimic (miR-661 bio; 100 nM) or biotinylated miR-control (miR-C bio), and harvested 48 hours later for pulldown analysis (see Materials and Methods). Fold enrichment with miR-661 relative to miR-C is shown for each indicated

mRNA. Values represent the average +/- SD from 5 independent experiments. P-Values for enrichment: Mdm4=0.03, Mdm2=0.002, beta actin=0.3; one-tailed Student's t-test). (B) Extracts of cells processed as in (A) were subjected to Western blot analysis with the indicated antibodies to validate the ability of miR-661 bio to downregulate Mdm2 and Mdm4.

4.2.3 miR-661 augments p53 functionality

Mdm2 and Mdm4 are both negative regulators of p53. Therefore, downregulation of Mdm2 and Mdm4 by miR-661 is expected to increase p53 functionality. One predicted manifestation is transcriptional activation of p53 target genes. I therefore monitored the impact of miR-661 overexpression on the endogenous levels of several such transcripts. Indeed, miR-661 overexpression significantly increased the amount of p21, *Btg2*, *CD95* and *Wig1* mRNAs, all transcriptional targets of p53 (Figure 8A).

4.2.4 miR-661 causes p53-dependent cell cycle arrest and inhibits cell proliferation

To investigate the biological impact of p53 activation following miR-661 overexpression, I next examined the effect of this miRNA on the cell cycle. As seen in Figure 8B, transfection of MCF7 cells with miR-661 led to a significant decrease in the S phase fraction, monitored by BrdU incorporation. This effect was partially alleviated by knockdown of p53 (Figure 8B), and was reproduced with a different p53 siRNA (data not shown). Similar results were obtained in WTp53-expressing H460 and A549 cells (Figure 8B); in A549, both the inhibitory effect of miR-661 and its alleviation by p53 depletion were particularly pronounced. These data suggest that the cell cycle inhibitory effect of miR-661 is mediated by a combination of p53-dependent and p53-independent mechanisms. Remarkably, miR-661 did not affect cell cycle progression in ovarian carcinoma-derived OVCAR-3 cells (Figure 8C), which harbor a missense mutation in the p53 DNA binding domain (DBD), or in OVCAR-8 cells that harbor a six amino acid in-frame deletion within the DBD[132], and are thus expected to have lost WTp53 function (Figure 8C).

Consistent with its inhibitory effect on cell cycle progression in MCF7 cells, miR-661 overexpression also led to a reduction in long-term colony formation capacity (Figure 8D). In conclusion, miR-661 overexpression can promote cell cycle arrest and reduce cell proliferation, at least partially through p53 activation.

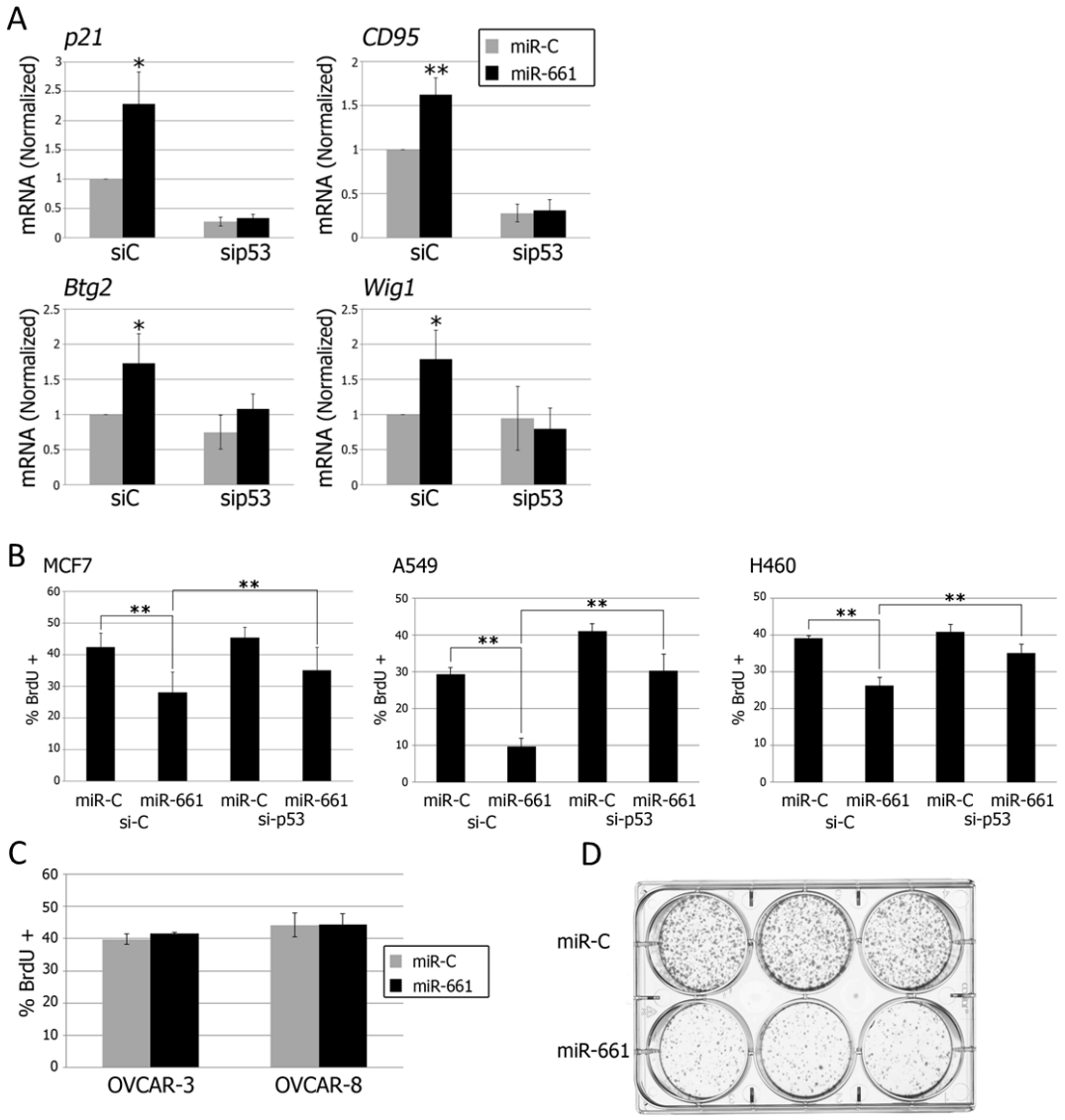


Figure 8: miR-661 upregulates p53 functionality and inhibits cell proliferation

(A) MCF7 cells were transfected as in Figure 6C. 48 hours later, RNA was extracted and subjected to qRT-PCR analysis of the indicated transcripts. Values were calculated as in Figure 6A. Values represent the average +/- SD from 3-4 independent experiments. P-Values for the difference between miR-661 and miR-C in the siC samples: p21=0.04, CD95=0.003, Btg2=0.05, Wig1=0.04; Student's t-test). (B) The indicated cell lines were transfected with miR-661 mimic (miR-661; 20 nM) or miR-control (miR-C) for 48h, followed by transfection of sip53 or siC. 24h later, cells were subjected to BrdU incorporation analysis as described in Materials and Methods. The percentage of BrdU-positive cells is shown. Values represent the average +/- SD from 3-4 independent experiments. P-Values for the indicated differences: MCF7 siC miR-661 vs. siC miRC=0.002, MCF7 miR-661 siC vs. miR-661 sip53=0.003, A549 siC miR-661 vs. siC miRC=0.002, A549 miR-661 siC vs. miR-661 sip53=0.002, H460 siC miR-661 vs. siC miRC=0.01, H460 miR-661 siC vs. miR-661 sip53=0.005.

(C) OVCAR-8 and OVCAR-3 cells were transfected with miR-661 mimic (miR-661; 20 nM) or miR-control (miR-C) for 48h, and subjected to BrdU incorporation analysis as in (B). (D) MCF7 cells were transfected with 20 nM miR-661 mimic or miR-control. 24h later cells were harvested and counted.

Equal cell numbers were seeded for a colony formation assay. 8 days later cell colonies were fixed, stained and photographed.

4.2.5 High miR-661 expression correlates with good prognosis in breast cancer

Overexpression of miR-661 activates p53 and exerts antiproliferative effects in MCF7 cells. If this also holds true for actual tumors, one might predict that higher miR-661 expression may restrict tumor growth and aggressiveness, at least in cancer types that share similar features with MCF7. MCF7 are derived from an estrogen receptor positive (ER+) breast cancer; therefore, I used the MIRUMIR tool[133] to query the prognostic value of miR-661 expression levels in patients with high-risk ER+ breast cancers, based on published data[130]. When I performed a Kaplan–Meier plot, I could show that patients with high miR-661 expression had a better survival probability than those with low miR-661 expression (Figure 9A, p-value=0.0002). Of note, ER+ breast tumors have a very low rate of p53 mutations, and therefore mostly express WTp53[134]. This observation is consistent with my *in vitro* findings, and suggests that reduced miR-661 expression may contribute to cancer aggressiveness, and possibly to therapy resistance, by attenuating p53 functionality in the tumor cells.

4.2.6 The miR-661 locus is preferentially amplified in tumors with mutant p53 and miR-661 promotes migration of cells from such tumors

My data suggests that miR-661 may be considered a putative tumor suppressor, since it induces antiproliferative effects, partly through augmentation of p53 activity. Surprisingly, analysis of genome-wide *miR-661* locus alterations employing the cBio portal[128] revealed that this locus actually tends to be amplified in a variety of cancers, including ovarian serous cystadenocarcinoma (~25% of cases) and invasive breast carcinoma (~7%) (Figure 9B). This might seem in disagreement with the proposed tumor suppressive effects of miR-661; however, further analysis revealed that while only a minority of tumors included in this dataset carried *TP53* gene mutations (overall p53 mutation frequency in the entire set of tumors = 44%), in most tumors with miR-661 amplifications the *TP53* gene was actually mutated (Figure 9B). Remarkably, miR-661 amplification is particularly frequent in ovarian serous cystadenocarcinoma (Figure 9B), a tumor type with an exceptionally high rate of *TP53* gene mutations[135,136]. Hence, miR-661 amplifications appear to be largely

avoided in tumors that retain WTp53, consistent with our prediction that, by boosting p53 functionality, such amplifications may interfere with tumor progression. Conversely, in tumors harboring p53 mutations, excess miR-661 may potentially become advantageous, favoring amplification of this locus.

Cancer-associated p53 mutations can endow the mutant p53 with cancer-promoting gain-of-function (GOF) activities[137,138]. Hence, in tumors harboring such mutations, miR-661 amplification might be favorably selected because it may sometimes stabilize the mutant p53 protein and augment its GOF effects. This possibility is supported by an experiment where ovarian carcinoma-derived OVCAR-8 cells were transiently transfected with either miR-661 mimic or miR-661 inhibitor. OVCAR-8 cells carry a 6 nucleotide deletion within the p53 DBD and accumulate stable mutant p53 protein. As seen in Figure 9C, miR-661 overexpression led to a modest increase in mutant p53 levels. Moreover, miR-661 inhibition partially reduced p53 levels, suggesting that the endogenous miR-661 indeed contributes towards sustaining mutant p53 accumulation in those cells.

One distinctive GOF activity of mutant p53 is augmentation of growth factor-induced cancer cell migration[139]. Indeed, depletion of endogenous mutant p53 markedly reduced serum-induced OVCAR-8 cell migration, confirming that the mutant p53 of these cells harbors GOF activities (data not shown). Importantly, miR-661 mimic overexpression significantly promoted OVCAR-8 cell migration (P-Value = 0.02, Figure 9D); conversely, miR-661 inhibition led to a modest but significant reduction in the rate of migration (P-Value = 0.04, Figure 9E). Hence, miR-661 can augment the migration of cells harboring mutant p53. Overall, these findings are consistent with the observed amplification of miR-661 in serous ovarian cancer, and suggest that such amplification might contribute to ovarian cancer progression partly through increasing mutant p53 levels.

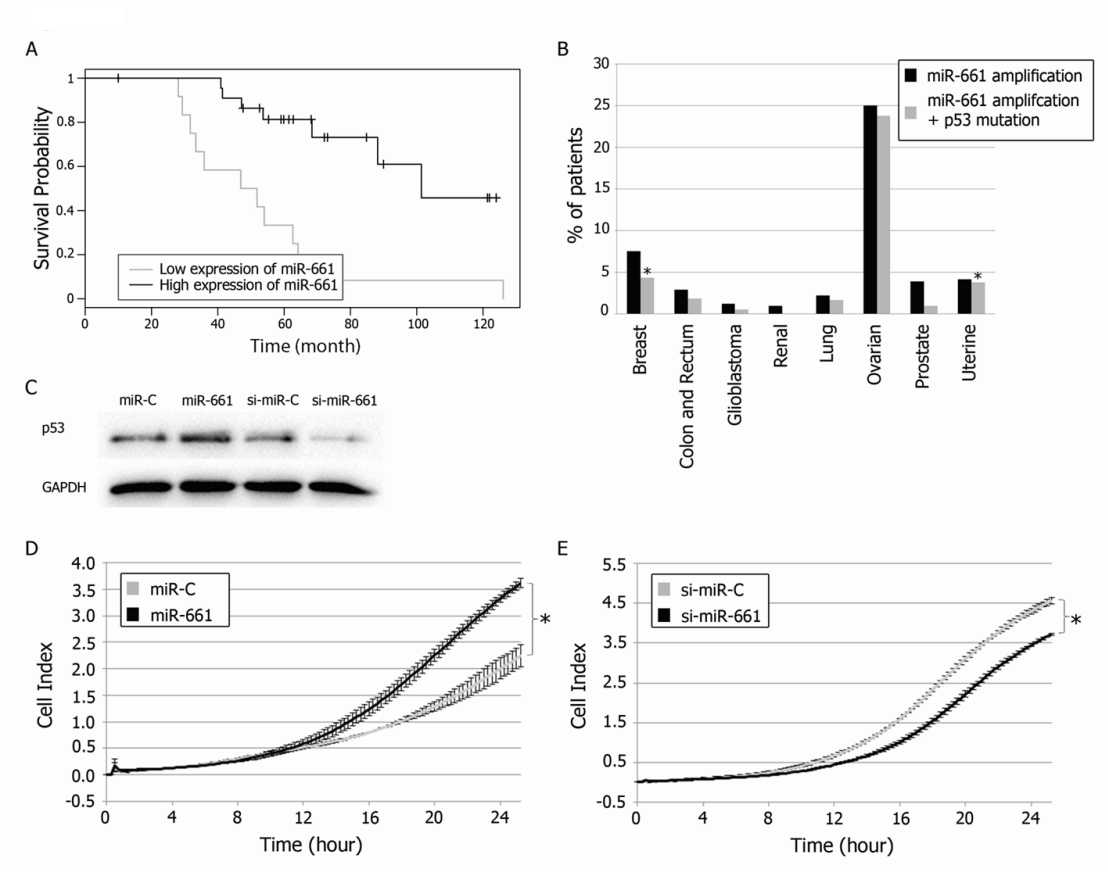


Figure 9: miR-661 in cancer patients

(A) Kaplan-Meier survival curves for patients with ER+ breast cancer expressing different miR-661 levels were calculated using the MIRUMIR tool. (B) For each cancer type the overall percentage of patients with miR-661 genomic amplification, as well as the percentage of patients having both miR-661 amplification and p53 mutation, was calculated using the cBio portal. P-values for non-random association between p53 mutations and miR-661 amplification: Breast invasive carcinoma=0.015, Colon and Rectum Adenocarcinoma=0.7, Glioblastoma Multiformae=0.1, Renal Clear Cell Carcinoma=1, Lung Squamous Cell Carcinoma=0.95, Ovarian Serous Cystadenocarcinoma=0.2, Prostate Adenocarcinoma=0.2, Uterine Corpus Endometrioid Carcinoma=6e-5; Hyper-geometric distribution). (C) OVCAR-8 cells were transfected with 20 nM miR-661 or miR-control (miR-C), or with 100 nM miR-661 inhibitor (si-miR-661) or miR-control inhibitor (si-miR-C) and harvested 48 hours later for Western blot analysis with the indicated antibodies. (D) OVCAR-8 cells were transfected with 20 nM miR-C or miR-661 for 48 h, and then subjected to a real-time migration analysis as described in Materials and Methods. All experiments were conducted in three biological replicates. Representative data from one of the replicates is shown. A t-test was performed for the last time point of all three replicates, revealing significant (P-Value = 0.02) differences in the means of the two populations. (E) OVCAR-8 cells were transfected with 100 nM si-miR-C or si-miR-661 for 48 h and analyzed as in (D). P-Value = 0.04.

5. 3’ UTR shortening potentiates microRNA-based repression of pro-differentiation genes in proliferating human cells

Alternative polyadenylation (APA) occurs in most genes under different conditions and in different cell types. It was previously shown that this process can eliminate miRNA regulation and enable proto-oncogenes to be upregulated in proliferating states and in cancer. Here I found that APA can also potentiate miRNA binding sites positioned just upstream to the APA site. This process occurs mostly in pro-differentiation genes, increasing their potential targeting by pro-proliferation miRNAs.

PLoS Genet. 2016 Feb 23;12(2):e1005879. doi: 10.1371/journal.pgen.1005879. eCollection 2016.

3'UTR Shortening Potentiates MicroRNA-Based Repression of Pro-differentiation Genes in Proliferating Human Cells.

Hoffman Y[1,2], Bublik DR[2], Ugalde AP[3], Elkon R[3], Biniashvili T[1], Agami R[3], Oren M[2], Pilpel Y[1].

Author information

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.
[2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.
[3]Division of Biological Stress Response, The Netherlands Cancer Institute, Amsterdam, The Netherlands.

Abstract

Most mammalian genes often feature alternative polyadenylation (APA) sites and hence diverse 3'UTR lengths. Proliferating cells were reported to favor APA sites that result in shorter 3'UTRs. One consequence of such shortening is escape of mRNAs from targeting by microRNAs (miRNAs) whose binding sites are eliminated. Such a mechanism might provide proliferation-related genes with an expression gain during normal or cancerous proliferation. Notably, miRNA sites tend to be more active when located near both ends of the 3'UTR compared to those located more centrally. Accordingly, miRNA sites located near the center of the full 3'UTR might become more active upon 3'UTR shortening. To address this conjecture we performed 3' sequencing to determine the 3' ends of all human UTRs in several cell lines. Remarkably, we found that conserved miRNA binding sites are preferentially enriched immediately upstream to APA sites, and this enrichment is more prominent in pro-differentiation/anti-proliferative genes. Binding sites of the miR17-92 cluster, upregulated in rapidly proliferating cells, are particularly enriched just upstream to APA sites, presumably conferring stronger inhibitory activity upon shortening. Thus 3'UTR shortening appears not only to enable escape from inhibition of growth promoting genes but also to potentiate repression of anti-proliferative genes.

PMID: 26908102 PMCID: PMC4764332 DOI: 10.1371/journal.pgen.1005879

5.1 Materials and Methods

3'Seq and miRNA array: WI-38 cells were grown in 37° in MEM supplemented with 10% non-heat-inactivated fetal bovine serum (Sigma), pen-strep, sodium pyruvate, L-glutamine solution (Beit HaEmek). RNA was extracted using Nucleospin miRNA kit (Macherey-Nagel), according to manufacturer's instructions.

microRNA array analysis was done in duplicates, using the miRNA Complete Labeling and Hyb Kit (Agilent, 5190-0456) according to the manufacturer's instructions. Briefly, for each sample 100ng RNA was dephosphorylated, denatured, labeled with Cyanine 3-pCp and purified using Micro Bio-Spin 6 Columns. Hybridization was done for 20h with Agilent SurePrint G3 Unrestricted miRNA 8x60K (Release 19.0) arrays. Arrays were scanned using an Agilent DNA microarray scanner, and analyzed using the AgiMicroRna package in R[140] with the RMA algorithm. The heat map was generated with Matlab Clustergram function for the mean fold change of each sample vs. control (mean of the duplicate arrays).

The 3'Seq protocol is based on Jenal et al.[91] and incorporates additional modifications described in Martin et al[46]. Basically, 25 μg of total RNA were heat-fragmented for 12 minutes in 1x Fragmentation Buffer (Ambion) at 70°C to generate RNA fragments of ~100 nucleotides. Next, the 3'end poly(A) RNA fragments were selected using the Oligotex mRNA Kit (QIAGEN) and RNA was end-repair with T4 polynucleotide kinase for 45 minutes at 37ºC following manufacturer's instructions. Afterwards, RNA 3'ends were blocked for ligation by incubation with 1mM Cordycepin 5′-triphosphate (Sigma) and 10U of polyA polymerase (PAP, NEB), in 1x PAP buffer for 30 minutes at 37ºC. Finally, a P7 RNA adapter (5'-CAAGCAGAAGACGGCAUACGAGAU-3') was ligated to the 5'end using 2U of T4 RNA ligase I and 2.5uM of RNA adapter, for 4h at room temperature. Between each step, RNA was purified using Agencourt RNAClean XP magnetic beads (Beckman Coulter) following the manufacturer's instructions. At this point, RNA fragments were converted to cDNA employing the Superscript III RT kit (Life Technologies) and an anchored oligo(d)T stem loop primer containing a barcoded Illumina adaptor as in Martin et al. Next, cDNA was purified twice with Agentcourt AMPure XP magnetic beads (Beckman Coulter) using a ratio 1.5:1 beads:sample. To generate the final 3'Seq library, the cDNA with the correct adaptor sequences was

enriched/amplified using Phusion DNA polymerase (Life Technologies) and primers P7 and Illumina_Truseq, for 12 cycles following manufacture's recommendations. Finally, the 3'seq library was size selected with AMPure XP magnetic beads by two rounds of purification with a ratio 1:1 beads:sample, before being sequenced on an Illumina HiSeq2000 system.

In our protocol of profiling transcript 3' ends, sequenced reads start with a barcode for sample multiplexing which is followed by six Ts whose end marks the precise location where the poly(A) tail starts. These six Ts therefore allow the mapping of poly(A) cleavage sites (CSs) with a nucleotide resolution. After trimming the barcode and six Ts, reads were aligned to the human genome (hg19) using TopHat[141]. Up to two mismatches were allowed in the reads' seed region (the first 28 nt). As CS location often fluctuates around a major site, we merged reads from all samples and identified "read runs'' (that is, genomic intervals that are "tiled" by multiple reads where the distance between the start of consecutive ones is below 10 nt). We considered the local maxima of these runs as the CS locations. We required a spacing of at least 50 nt between consecutive CSs. (In case of lower spacing between CSs, the one supported by a higher number of reads was chosen). Only CSs supported by at least 10 reads (at the location of the CS run maximum) were considered in subsequent analyses. The median length of the runs was 11 nt. Overall, 41,972 CSs were detected in our dataset. Priming of the oligo-dT primer to genomic regions that are A-rich ("internal priming") could lead to false call of CSs. To reduce the rate of such false calls we extracted genomic sequences of 50 nt centered at the location of the putative CSs, and filtered out CSs that contained in that region a stretch of 10 nt of which at least 8 were As and the rest were Gs. 4,307 suspected CSs were filtered out.

Prediction of miRNA binding sites and conservation analysis: miRNA binding sites were defined as perfect 7-mers, which are reverse complement to the seed of the miRNAs, for all human and mouse miRNAs listed in miRBase release 17[104]. Conserved binding sites were taken from TargetScan release 6.2[113].

P_{CT} and Context++ scores: P_{CT} scores of all miRNA binding sites of type "7mer-m8" (perfect 7mer) for the conserved miRNA families were taken from TargetScan release 6.2[113]. Context++ scores of all miRNA binding sites of type "7mer-m8" were taken from TargetScan release 7[7]. Only miRNA binding sites from genes with at least 500 bases around the APA sites were taken into the analysis.

Analysis of genes with APA sites: All analyses were done only for genes whose accession number was included in TargetScan 3'UTRs list (as defined in their website). A cleavage site was assigned to a gene if it was included in the coordinates of its 3'UTR (and 20 bases further, after the 3' end of the 3'UTR). A gene was considered to have an APA site only if it has at least two cleavage sites assigned to its 3'UTR. In all analyses, the APA site that we took into account was the 5' most in the 3'UTR. To compute the statistical significance of the main signal – enrichment of conserved miRNA binding sites at particular distances from the APA site, we created a randomized null model. In this null model each binding site's location was recorded and random position of an APA site was drawn from the full length 3' UTR for each gene (omitting the first and last 1000 bases in order to allow inspection of that vicinity around the randomly chosen location). The distance from the binding site and the randomized APA location was computed and the procedure was repeated 10,000 times. This yielded a distribution of distances as shown in all plots, once for conserved binding sites, and once for all binding sites. * indicates p-value < 0.05 for the null hypothesis that for a specific distance from the APA site, the number of conserved binding site is similar as in a random APA site, or higher.

Conservation profiles around conserved miRNA binding sites: PhastCons and PhyloP scores of each base in the genome (hg19) were taken from UCSC[142,143]. The profile around APA of genes was for genes with at least 1000 bases from each side of the APA sites. The profile around conserved miRNA binding sites was for all miRNA binding sites located in the 300 bases 5' to the APA site. The sites were aligned and the mean conservation profile was calculated.

Codon usage and miRNAs analysis: I defined the "Pro-Proliferation" and "Pro-Differentiation" gene sets as follows: I began with the Gene Ontology sets termed "M-phase of cell cycle" and "Pattern Specification", two gene sets that were recently shown[144] to serve as archetypical proliferation and differentiation genes, with distinct codon usage. To augment the number of genes belonging to each of the two sets we searched the entire genome for additional genes whose codon usage was highly similar to either of the two groups, thus expending the two sets from 92 and 82 genes originally to 229 and 136. After computing the miRNA binding site distribution around APA sites for the genes in each of the sets we estimated a p-value on the difference between miRNA density at each distance from APA sites as follows: we repeated 10,000 on randomly partitioning the genes with high correlation to the two

groups into two groups, one with 229 genes and one with 136 genes for the codon usage expansion threshold 0.75 correlation. In each such random partition we recorded, at each location relative to the APA, the fraction of conserved miRNA binding sites. The p-value was estimated as the fraction out of the 10,000 repetitions in which the real partition into "Pro-Proliferation" and "Pro-Differentiation" resulted in a difference in binding sites count. * indicates one-sided p-value < 0.05 for the null hypothesis that for two groups of these sizes, the difference in number of conserved binding sites is as for the two original groups or higher. The miR-17-92 binding sites analysis was similar to the codon usage. Here too randomization was done 10,000 taking a random group of genes in the same size of the genes with binding sites for miR-17-92 miRNAs (104), and computing for each distance from the APA site the difference in the number of conserved binding sites for the random group and for all genes. * indicates one-sided p-value < 0.05 for the null hypothesis that for a random group of genes in the same size as the original one, the difference between the number of conserved binding sites between this group and all genes is as good as for the original group or higher.

Folding prediction: Folding prediction was done using the RNAfold algorithm in Vienna Package 2.1.9[145] with default parameters. The contact matrices of all sequences in each analysis were summed and averaged using perl, and the plots were generated with Matlab. The p-value analysis for the difference in the averaged contact matrices of two groups was done by 10,000 random separations of the original sequences to two groups in the same size as the original groups, and comparing the difference in the random separation to the original difference between the two groups. FDR was calculated in Matlab.

5.2 Results

5.2.1 Conserved miRNA binding sites are enriched immediately 5' to APA sites

In order for an APA event to potentiate targeting by miRNAs, potentially functional sites should exist 5' to the APA site, at a distance from the APA that is comparable to that typically seen between conserved miRNA sites and the canonical full-length 3' UTR ends. I have previously shown this region to be ~250 nucleotides from either end of the 3'UTR[131].

To identify transcriptome-wide APA sites, WI-38 human embryonic lung fibroblasts and their immortalized derivatives obtained through sequential serial transfers towards increased proliferation and transformation[146,147] were subjected to 3' sequencing and analysis[148]; this was done in collaboration with the lab of R.Agami at the NKI, Amsterdam. The analysis included primary fibroblasts ("Control"), slow growers (early passage after immortalization), fast growers (extensive passaging after immortalization) and fast growers transformed by constitutively activated mutant H-RasV12 ("Ras"). 5765 genes were found to have at least one APA site in at least one of the cell types in this experimental system. Somewhat unexpectedly, I did not detect significant differences in the overall extent of global shortening between the different cell types. Remarkably though, when I aligned all genes with at least one APA event according to their most proximal APA site, a significant enrichment of conserved binding sites was observed within the 300 bases immediately upstream to the APA site (Figure 10A for control cells, Figure 10B for all time points). A very similar picture emerged from the analysis of previously published 3'seq data[91,148], obtained in different cell lines (Figure 10C-E). A trivial reason for the similar signals in the different datasets could be that in all those datasets the same genes undergo the same APA events. However, the same pattern was still retained also when I performed a similar analysis only on APA-positive genes that differ between pairs of cell lines (examples in Figure 10F-H). Moreover, analysis of 3'seq data from mouse muscle tissue[91] revealed a similar peak of conserved miRNA binding sites upstream to the APA sites (Figure 10I). In the mouse tissue fewer genes were found to undergo APA, an observation which might reflect the highly differentiated state of the cells. Hence,

in multiple cell types and in two mammals, distinct APA sites reside preferentially closely downstream to conserved miRNA binding sites, effectively repositioning such centrally-located miRNA binding sites and placing them in proximity to the 3' end of the shortened transcript. In support of the emerging notion, when I compared the distribution of conserved miRNA binding sites between genes possessing at least one APA site in WI38 cells and those without APA sites I found (Figure 10J) that genes with APA sites tend to harbor fewer conserved binding sites near the distal end of their full length 3'UTR, relative to those without an APA (p-value=6.5e-279, Student's T-test). Conversely, genes with APA sites are relatively enriched in conserved miRNA binding sites within the proximal half of their 3'UTR (Figure 10J). This observation is intriguing: if APA merely serves to eliminate miRNA binding sites residing near the 3' end of the full length mRNA, one would expect APA-positive genes to be more enriched for functional miRNA binding sites near that end, providing them with an efficient on/off switch controlled by APA. The fact that the opposite trend is actually observed strongly suggests that the interplay between APA and miRNAs may allow regulation that is richer than mere binding sites elimination. Specifically, this may serve as further indication that the shorter 3' end, positioned immediately upstream to the APA site, can dynamically potentiate new miRNA binding sites as they become positioned closer to the 3' end of the shorter transcript.

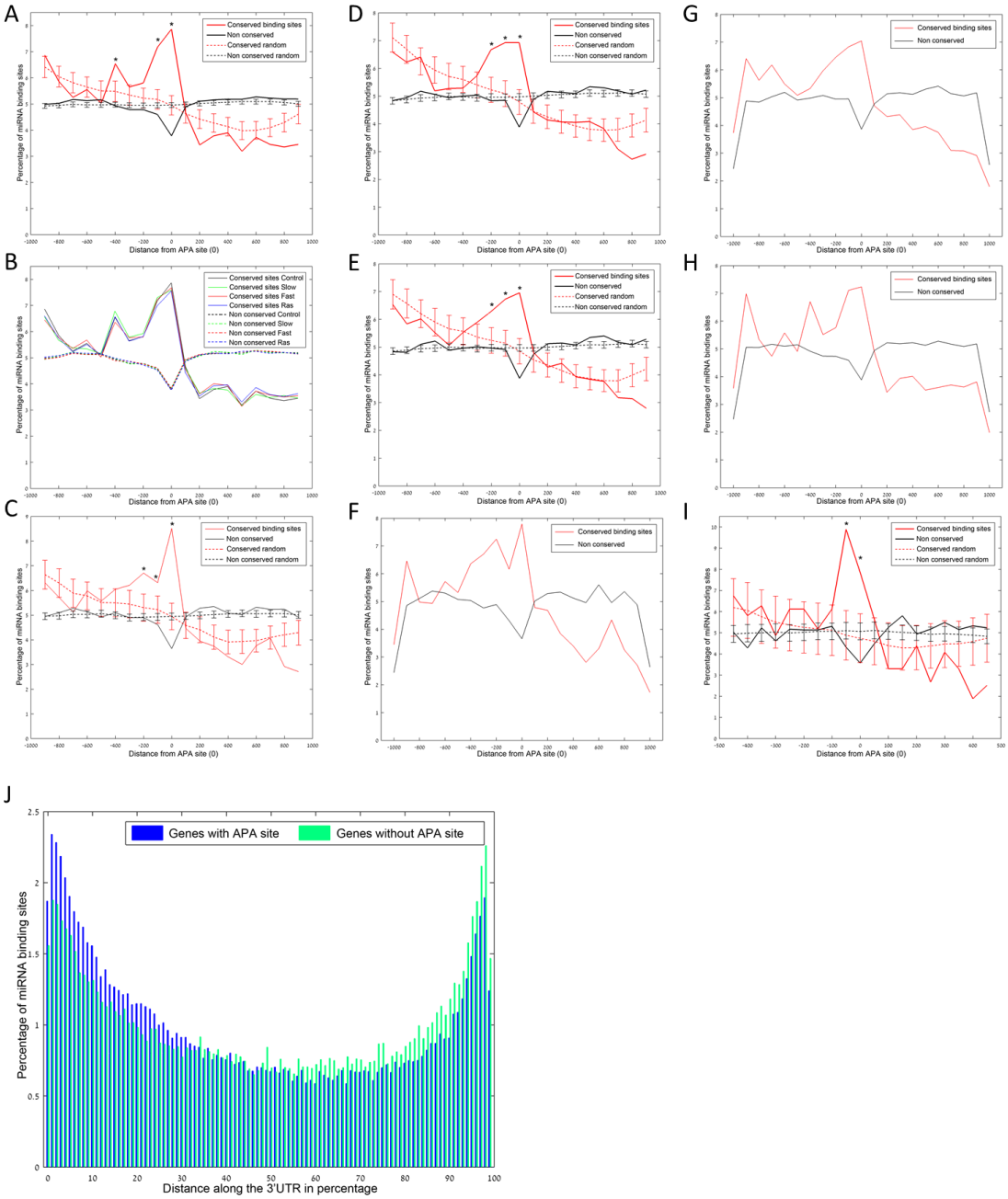


Figure 10: Conserved miRNA binding sites are enriched immediately 5' to APA sites

Conserved and non-conserved miRNA binding sites around the APA site (point 0), for all genes with at least 1000 bases of 3' UTR sequence from each side of the APA site, in different cell lines: WI38 control cells (A), WI38 all time points (B), U2OS (C), BJ (D), MCF10A (E) and Mouse Muscle Tissue (Genes with at least 5000 bases 3' UTR sequences from each side of the APA site) (I), or for genes appearing in one cell line and not in the other: U2OS not in BJ (F), BJ not in MCF10A (G), and WI38 not in BJ (H). (J) Conserved miRNA binding sites along the 3'UTR in percentage, for all genes with 3'UTR length of at least 1000 3' UTR bases, with and without APA site in WI38 cells. * indicates p-value<0.05.

5.2.2 miRNA binding sites located 5' to APA sites are probably selected for miRNA targeting

Analysis of the conservation state of each miRNA binding site by itself is an accepted indication for its functionality[3,112,113]. However, such conservation might be due to other attributes, for example another functional feature of the 3'UTR residing in this location. I therefore asked whether the high conservation of the miRNA sites located upstream to APA sites can be attributed to conservation of their neighborhood, or is preferentially targeting the miRNA sites? To address this question I looked at the conservation profile around APA sites, for genes with and without conserved miRNA binding sites in the 300 bases 5' to the APA sites. As can be seen in Figure 11A, the profile of the genes without conserved sites 5 to APA sites is lower specifically in this area, indicating that the high conservation present for the other genes might be due to the presence of miRNA targeting, and not merely APA sites. I also looked at the conservation profile of all conserved miRNA binding sites located in the 300 bases 5' to APA sites. For that I used PhastCons and PyhloP[142,143] as provided by the UCSC browser. I observed a sharp peak of conservation exactly overlapping the 7 bases of the binding sites, whereas its surroundings are significantly less conserved (Figure 11B), strongly arguing in favor of a selective pressure to conserve specifically the conserved miRNA target sites. Another measurement of site conservation is the P_{CT} score, which controls for the 3'UTR surroundings, dinucleotide conservation and other parameters unrelated to miRNA functionality[149]. Importantly, this score allows to assess the extent to which conservation of a site is likely to be due to miRNA functionality. We compared the distribution of P_{CT} scores for miRNA binding sites 300 bases 5' and 3' to APA sites, and found that the scores are significantly higher for the sites located just 5' to the APA sits (Figure 11C, p-value=2e-7, Student's T-test). Moreover, I employed the Context++ scoring system of miRNA binding sites, which takes into account many additional parameters of each site and its surroundings beyond mere conservation, and provides a score for the probability that this site is indeed functional[7]. I compared the scores of the binding sites 300 bases 5' and 3' to APA sites, for conserved and non-conserved sites. For both groups, the scores for sites located 5' to APA sites were significantly lower (hence indicating higher functionality) (Figures 11D,E, p-value=2e-6, 7e-165, Student T-test). Together with

the P_{CT} score analysis, this argues that these sites are more likely to be indeed functional.

In sum, the above analyses strongly suggest that the miRNA binding sites located just 5' to APA sites are functional beyond sequence conservation. Since the Context++ scores of non-conserved miRNA binding sites were also better for the ones that are positioned just 5' to APA sites, I can conclude that even the non-conserved binding sites are probably more functional for miRNA targeting when located in that region.

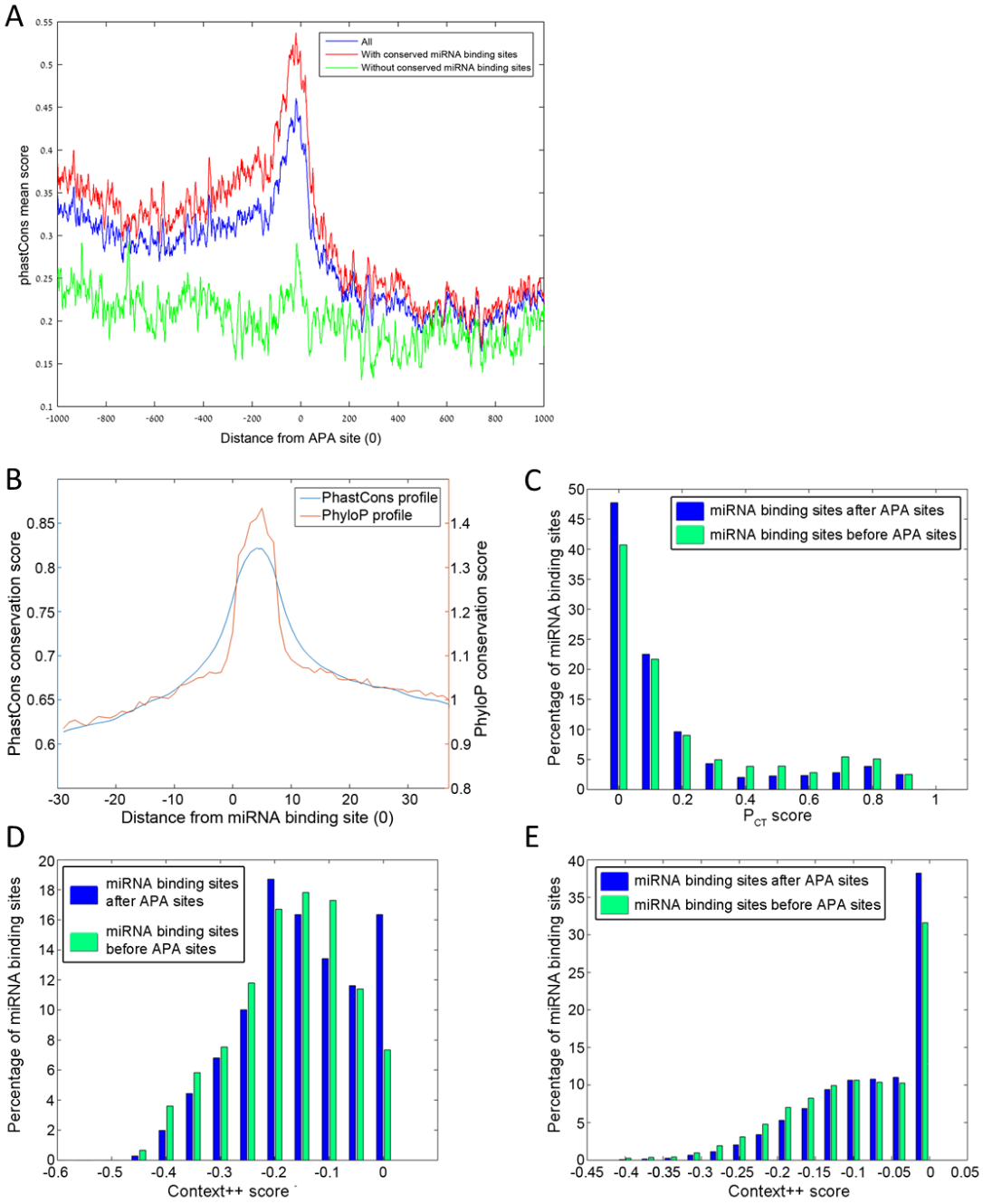


Figure 11: miRNA sites located 5' to APA sites are probable functional sites for miRNA targeting
(A) PhastCons conservation profile around APA sites of genes with at least 1000 bases from each side of the APA site. The genes are divided into those with, or without, conserved miRNA binding sites in the 300 bases 5' to the APA site. (B) Conservation profile of 30 bases around conserved miRNA binding sites which are located in the 300 bases 5' to APA sites. Two conservation scoring systems are displayed – PhastCons and PyhloP. (C,D,E) P_{CT} conservation scores (C) and Context++ scores (D,E) of miRNA binding sites located 300 bases 5' or 3' (before and after) APA sites, for genes with at least 500 nucleotides from each side of the APA site. For the context++ scores the miRNA binding sites are divided to conserved (D) and non-conserved (E).

5.2.3 Embryonic pattern specification genes and targets of the proliferation microRNA cluster are enriched for conserved sites immediately upstream to APA sites

In non-proliferating cells, most genes express mainly the full length version of their 3'UTR. However, in proliferating cells and particularly in cancer, increased usage of APA sites has been shown to enable miRNA binding sites elimination from the 3'UTRs of proliferation-associated genes[87,93]. By the same rationale, I predicted that if our above observations are physiologically relevant, then genes becoming more susceptible to miRNA regulation (and hence more effectively repressed) owing to APA would tend to be those that should be preferentially downregulated during proliferation and in cancer. To address this prediction, I initially compared two gene sets: those at the core of the cell cycle machinery, and those involved in patterning of the embryo during development. I chose these two gene sets as they represent two opposing classes of archetypical proliferation and differentiation genes[144]. Yet, these gene sets are relatively small and for only a portion of them we detected APA events. I therefore sought to expand these gene sets to include functionally related genes and thus gain further statistical power. To that end, I expanded each gene set to include additional genes either by similar codon usage. I then analyzed the miRNA binding site landscape around the APA sites of the two gene sets. Remarkably, I observed a significant difference between the two groups of genes: while the proliferation-related genes ("Pro-Prolif.") display only modest enrichment of conserved miRNA binding sites immediately upstream to the APA site, the differentiation-related genes ("Pro-Diff.") show a markedly elevated abundance of conserved miRNA binding sites in the corresponding region (Figure 12A). This strongly suggests that differentiation-related genes are more prone than proliferation-related genes to regulation by miRNA binding site potentiation via APA. This is in line with the documented increased APA usage during proliferation and cancer, when differentiation-related genes are expected to be downregulated.

These results suggest that, in addition to its documented ability to alleviate miRNA-mediated repression of proliferation genes, 3' UTR shortening is also used to potentiate preferentially the repression of differentiation genes.

5.2.4 Binding sites of a pro-proliferation miRNA cluster may be potentiated by APA

To obtain clues about the miRNAs that bind binding sites potentiated by APA, I performed comparative miRNA microarray analysis on WI-38 cells and their progressively transformed derivatives[146,147]. I found that most of the miRNAs that were upregulated particularly in the highly proliferative stages (fast growers and Ras-transformed) belong to the miR-17-92 cluster (Figure 12B). Indeed miR-17-92 is a well-studied proliferation-associated cluster[150], which includes 6 miRNAs with 4 different binding sites sequences. Notably, in comparison to all miRNAs on the array, conserved sites for members of the miR-17-92 cluster are significantly enriched immediately 5' to APA sites (Figure 12C). Furthermore, while for all miRNAs in the genome I saw a marked increase in the abundance of conserved binding sites near the 3' end of the full length 3'UTR, conserved binding sites of the miR-17-92 cluster are relatively less enriched in that region (Figure 12D). Thus, as compared to the bulk of the cellular miRNAs, miR-17-92 cluster members preferentially have binding sites that are located upstream to APA sites rather than near the distal end of the full length 3'UTR. APA is therefore expected to preferentially potentiate the repressive effects of those proliferation-associated miRNAs.

Overall, the above findings further support the conjecture that while APA enables proliferation-associated genes to escape miRNA regulation, it confers increased regulation upon pro-differentiation genes.

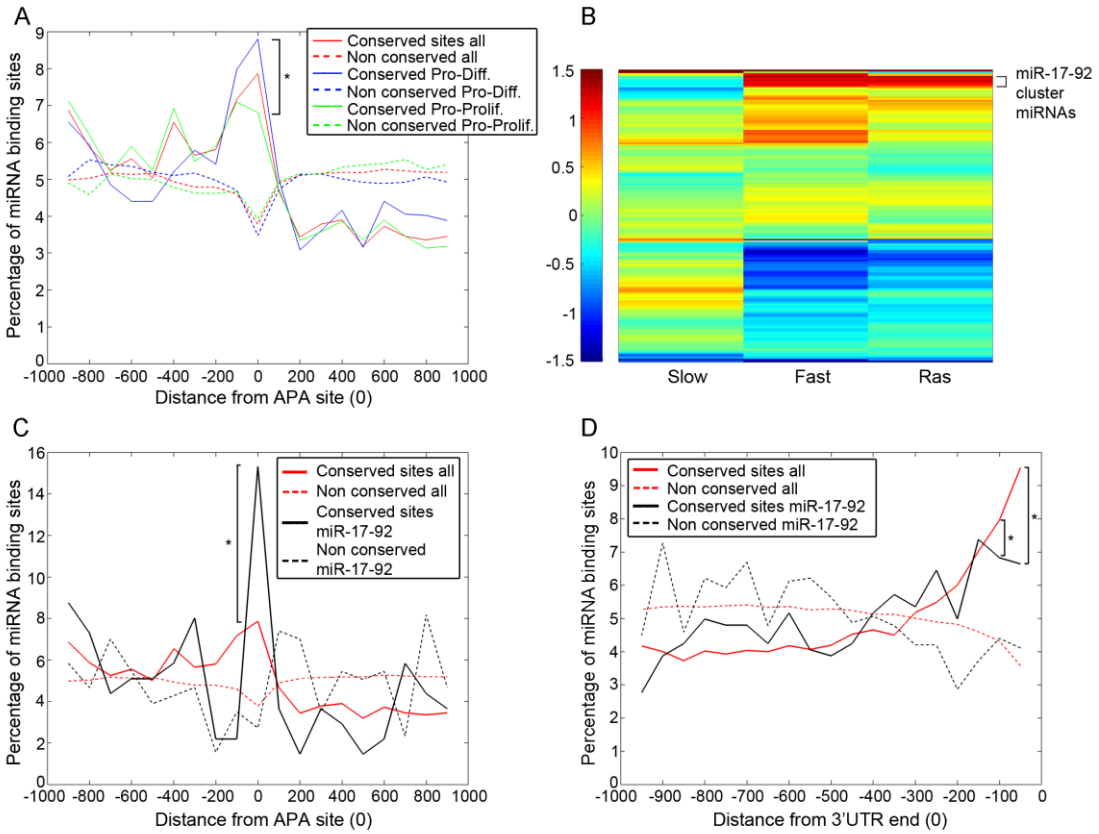


Figure 12: miRNAs and genes enriched for conserved binding sites upstream to APA sites

(A) Conserved and non-conserved miRNA binding sites for genes with APA site and at least 1000 3' UTR bases around it are divided in different groups according to codon usage correlation. (B) Heat map representing the mean fold change (log2 scale) of each miRNA in the miRNA array experiment from the control (primary cells) sample. (C) Conserved binding sites around the APA site for miRNAs in the miR-17-92 cluster and for all miRNAs. Only genes with at least 1000 3' UTR bases before and after the APA site were considered for the analysis. (D) Conserved binding sites upstream the long 3'UTR end for miRNAs in the miR-17-92 cluster and for all miRNAs. Only genes with at least 1000 bases 5' to the long 3'UTR and with APA site were considered for the analysis. * indicates p-value<0.05.

5.2.5 Tight secondary structure of the RNA near APA sites renders miRNA binding sites less accessible when the long UTR is used

If APA can indeed switch on miRNA site residing in close proximity 5' to the APA site, it is plausible that such binding sites might be masked fully or partially from the miRNA machinery when the APA site is not used and a longer form of the RNA is prevalent.

In order for a miRNA binding site to be functional, it should be in a relatively loose secondary structure context, rendering it accessible and energetically favorable for

miRNA binding[5,6]. Therefore, we computed the RNA secondary structure of the regions around the shortest APA site using the RNAFold algorithm[145], for all genes with an APA event in WI-38 cells. I reasoned that interactions between RNA stretches positioned on the two sides of the APA site might base pair with one another, thus masking potential miRNA binding sites just upstream to the APA. Such masking, if existing, would be relieved upon cleavage of the RNA at the APA site. Indeed, I found that when compared with randomly shuffled ensembles of these same UTR sequences (Figure 13A, right), sequences around APA sites have a relatively tight secondary structure (Figure 13A, left). Figure 13A left and right, respectively, show the fraction of sequences at which each pair of nucleotide positions in the region around the two sides of an APA are predicted to be paired in the folded structure in the real sequence (left) and in a randomized version of the sequence (right). Interactions consistent with a hairpin structure centered around the APA are predicted to occur often in the real structure, more than in the randomized sequence. I computed an empirical p-value for the hypothesis that each pair of nucleotides is more often engaged in a base-pairing interaction in the predicted folded structure of the real sequences as compared to their randomized shuffled forms. Although most differences did not pass an FDR of 10%, we still see a pattern in the significant locations with p-value<0.05. This observation is in line with Ding et al., who observed structural elements near the APA sites in Arabidopsis thaliana[151]. It is conceivable that such tight structures might serve primarily as one of the features that define an APA site irrespective of miRNA sites. Regardless, the outcome is that potential binding sites residing close to the APA site are predicted to be less accessible to the miRNA machinery. My model furthermore predicts that the structure around the APA should become looser when the APA site is used and the 3'UTR is shortened. Indeed when I computed the structure of the region upstream to the APA site when it becomes the end of the RNA, I found that the structure is now more open, and the folding energy of the sequence is less negative (Figure 13B,C, p-value=0.01, Student's T-test).

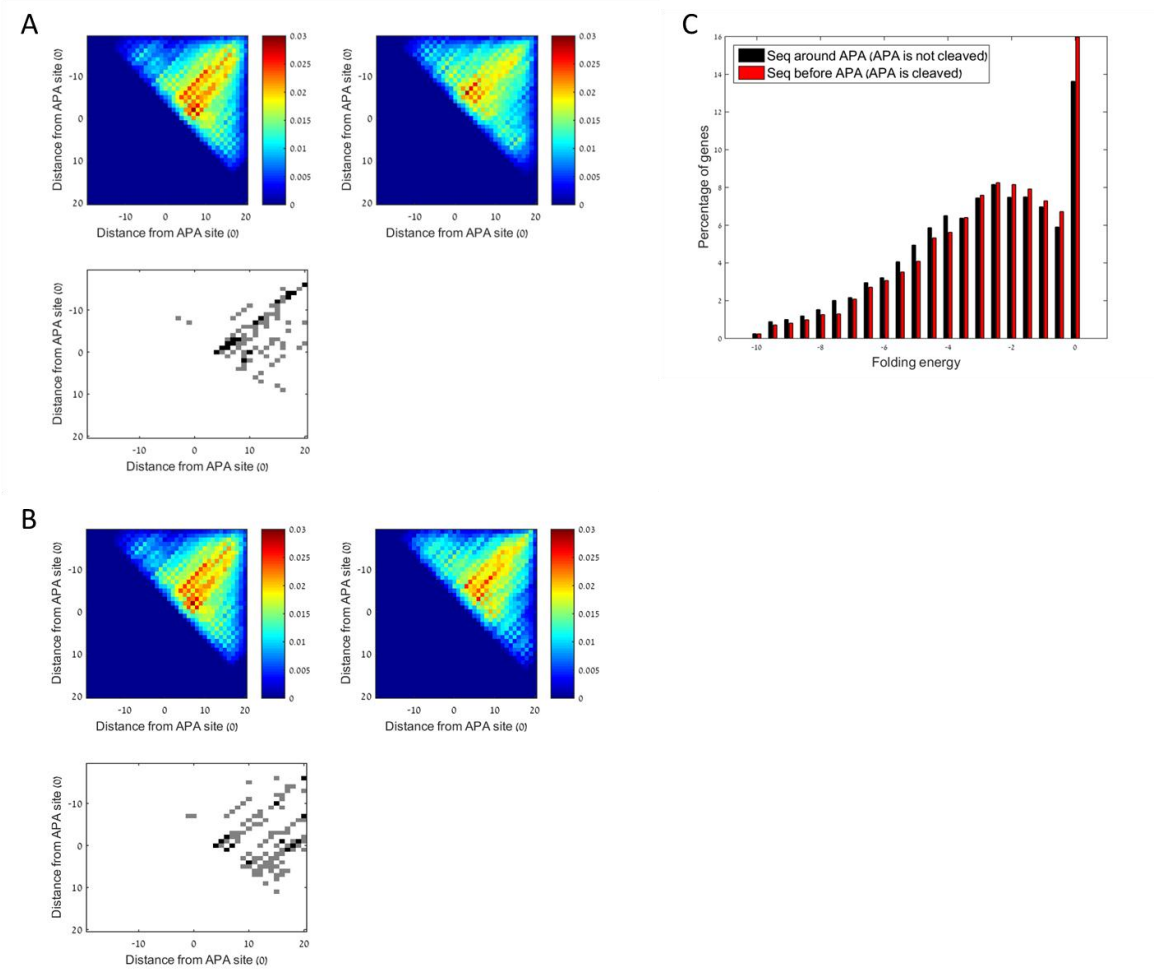


Figure 13: Secondary structure around APA sites

(A,B) Each dot in the heat map represents the average number of times these two bases are in contact, along all sequences in the analysis. The gray-black map represents the significance of each dot between the two heat maps in comparison, gray is p-value<0.05, black is q-value<0.1. (A) Compares all sequences with APA site vs. shuffled sequences. The APA is in point 20. (B) Compares all sequences with APA site where the APA is in the middle of the sequence (point 20), and all sequences where the APA is the end point (40). (C) Folding energy of the sequences in (B).

6. Discussion

miRNAs are small non-coding RNAs with a large impact on gene regulation. An important constraint for a miRNA to regulate a specific mRNA is that the mRNA should possess a binding site for the miRNA in its 3'UTR[1]. Such binding site for a miRNA is typically only 7-bases long[2], while the average length of a 3'UTR (in humans) is ~1000 bases[152]. Just by sequence similarity, typical genes are expected to contain hundreds or even thousands of potential binding sites for many miRNAs. However, the number of miRNAs that in fact bind and regulate each mRNA is usually much smaller[153]. Therefore, it is obvious that having a potential binding sequence for a miRNA within the 3'UTR of a transcript cannot be the only feature dictating whether a miRNA will in fact target a particular mRNA species.

A key question, which still remains partially open and has been investigated since the early days when miRNAs were discovered, is what else determines the binding of a miRNA to a specific mRNA[2].

During my PhD work I tried to answer this question from 3 different angles: can Alu elements spread miRNA regulation, and what constrains them from doing that indiscriminately? Is miR-661, a primate-specific miRNA predicted to target many Alu-embedded sites, a regulator of key molecules in the p53 network? Can alternative polyadenylation be an on-off switch to potentiate miRNA regulation, within the mRNA itself? While the first and third questions were addressed on a global scale, asking a question that is relevant to many genes, primarily using bioinformatics, the second question enabled me to zoom in to a specific miRNA-mRNA combination, and to carefully investigate their interaction and its consequences under different conditions, using mainly molecular cell biology tools.

The first angle for the main question (=what enables miRNA binding) was looking at Alu sequences, which are present in many transcripts' 3'UTRs, and are spread all over the human genome[21], and asking whether they can spread miRNA regulation together with their ability to jump and insert themselves into new locations within the genome[21]. I found that most potential miRNA sites within Alus are non-functional, and the miRNA machinery avoids binding to them. However, despite the strong indications that Alu-contained putative miRNA binding sites do not tend to affect gene expression, there are clearly cases where such binding sites within Alus can be functional. In particular, Smalheiser and Torvik described many mRNAs that contain

Alus in their 3'UTR, within which there are binding sites for dozens of miRNAs[41]. In their study most of the miRNAs suggested to target Alus were derived from the C19MC cluster[41], which is a primate-specific cluster that contains many Alu sequences that might have facilitated its expansion[154]. These miRNAs might have evolved in coordination with the Alu sequences, to create an effective targeting network. However, little is known about these miRNAs and their expression is undetectable in most developmental stages[155]. Lehnert et al. reported that there are a few miRNAs with more than 1000 predicted sites per megabase within Alu sequences, and proposed that such miRNAs protect against Alu transposition[40]. I found that although the potential regulatory effect of Alus is huge, their actual contribution to regulation of gene expression by the miRNA machinery might be limited. Clearly, this does not exclude other regulatory roles of the miRNA-Alu interplay, such as a role of miRNAs in maintaining genomic stability by the repression of transposable elements[156].

Despite the above general conclusion about the overall non-functionality of miRNA binding sites within Alus, I found that miR-661, a primate-specific miRNA that regulates two major components of the p53 network, Mdm2 and Mdm4, targets primarily sites located within Alu sequences in the 3'UTRs of those mRNAs. This is an example of a rare exception to this general rule, which might present an opportunity for the primate genome to acquire novel regulatory layers. One hypothesis for the existence of such rare example is the multiplicity of binding sites within each mRNA (3 in *Mdm2,* 9 in *Mdm4*), which may help overcome the low affinity of the interaction of each site by itself. Indeed, when I cloned specific binding sites into a luciferase-renilla reporter, each site by itself didn't exert any measurable impact. Such effect for multiplicity of miRNA binding sites was previously described[157], and one explanation may be the synergistic effect of such multiple sites[158].

The third angle in terms of what determines the functionality of a potential miRNA binding site is different than the previous ones. Most features that impact the functionality of a potential miRNA site are permanent and do not change (except for changes in the expression of the miRNA itself). However, alternative splicing adds another, dynamic dimension to the miRNA-mRNA interplay. I found that alternative polyadenylation (APA) can function as an on-off switch for the functionality of the miRNA binding site. When the APA site is not cleaved the binding site is less

functional, as it is positioned in the middle of the 3'UTR, which is a less favorable location for miRNA binding[3]. However, when the APA site is cleaved, the potential binding site positioned just prior to it becomes close to the new end of the short 3'UTR, and can now be visible to the miRNA machinery. Such switch may enable the mRNA itself to activate or shut down miRNA regulation, according to the changes in position in the cell cycle or due to external stress.

My work combined global bioinformatics analysis, together with molecular-based techniques, to detect single pairs of miRNA-mRNA interaction. An important aspect to discuss is the relevance of such global bioinformatics analysis to the miRNA world. In an ideal scenario, we could know all the pairs of miRNA-mRNA interactions. Clearly, this is not the case now, and bioinformatics and global analysis of miRNA target prediction and rules are primarily employed in order to direct scientists towards experimentally testable real functional mRNA targets for each miRNA, since without these algorithms the work is endless. However, if indeed such ideal case will happen one day, I still believe that global analysis is needed. First of all, because knowing the correct target set of each miRNA is not enough, one needs to take into account also the conditions and cell specificity of each targeting event, as in the case of alternative polyadenylation that can change the functionality of a binding site during the life time of a cell. Another important aspect of global and bioinformatics analysis in the miRNA world is interactions of networks. There is a small number of miRNAs that by themselves contribute dramatically to cellular events; however, for the large majority of miRNAs this is not the case. It is possible to think that miRNAs work in groups, influencing the same mRNA or the same group of mRNAs together, to lead to a particular cellular outcome. Here global bioinformatics analysis can contribute in revealing such networks and connected interactions between miRNAs.

Another similarity and common subject between the 3 sub-projects presented in this thesis is evolution. Alus represent a primate-specific genomic burden, and one of the solutions of the genome to avoid the spread miRNA regulation via Alus is an evolutionary one. Mutually exclusive territories (Alus are depleted from the 3'UTR ends, locations that are more responsive to the miRNA machinery[3]) imply an evolutionary mechanism, as it appears that Alu insertions near 3'UTR ends were selected against. One intriguing hypothesis is that Alus inserted near the ends of 3'UTRs might have forced important miRNA binding sites to move towards the

middle of the 3'UTR, where they have become less effective. Another scenario, supported by my findings, is that Alus near the ends might have introduced new miRNA binding sites at locations where such binding sites are likely to be highly functional, grossly disrupting 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 miRNA function.

miR-661 is a primate-specific miRNA, and its binding sites within *Mdm2* and *Mdm4* mRNA are mostly within Alus, which are also primate-specific. However, the p53 network is highly conserved and is present in all mammals and even in the Fugu fish[159]. While Alus represent an evolutionary burden for the genome in spreading miRNA regulation into genes where it may be deleterious, it can also present an opportunity for new primate-specific regulation. The case of miR-661 and Mdm2/Mdm4 is exactly the rare example where such regulation became beneficial instead of deleterious. Adding new layers of primate-specific regulation to such a conserved and important network such as the p53 pathway was enabled due to the ability of Alus to insert themselves into new genes in multiple copies, and the miR-661 regulation in this network adds new benefits and capabilities to it.

In the APA-miRNA story, evolution has an important role in the creation of such interplay. Was the APA site inserted before or after the insertion of a large number of potential binding sites for miRNAs in this area? The story can also be in the opposite direction – an APA site was created in a favorable location for other reasons, and then there was a selective pressure for the introduction of miRNA sites prior to it. Another interesting question regarding evolution is whether evolution inserted an APA into an existing long 3'UTR, or was actually a shorter UTR extended by adding a more distal polyA site, which gradually became more dominant? This may be the subject of further research.

Another interesting thread that passes through the three projects that are presented in this thesis is the time scale. We can look at each organism in two time scales: evolutionary and physiological. The evolutionary time scale is long, and relates to all organisms from the same species. It spans deep changes within the genome of the organism, including mutations, deletions, insertions and phenomenal changes, which can lead to the creation of a new organism. The physiological time scale is much shorter, and relates to each individual in the population. Most changes within this time scale are temporary, and non-heritable.

miRNA regulation is usually viewed in an evolutionary time scale, for example conservation of binding sites and miRNAs themselves. The Alu and the miR-661 projects deal with the evolutionary time scale, while the APA project is about changes that occur in a physiological time scale.

Potential miRNA binding sites within Alus exist from the first Alu insertion into the primates' genome. However, according to my findings their functionality was halted; or from a different point of view – they were not allowed to gain additional features that will make them functional (besides the correct binding site sequence). This process happened along the evolutionary time scale of primates. The miR-661 binding sites present an opposite example of an exception to the general conclusion of the Alu – miRNA study, as the binding sites of miR-661 in Mdm2 and Mdm4 within Alus did evolve towards functionality during primate evolution. It also represents an interesting regulation layer for the primates, which is not conserved and doesn't exist in other mammals. One of the most dominant features in miRNA target prediction algorithms is the conservation of the site. The case of miR-661 should be a warning sign not to rely too strongly on the conservation aspect, as important and interesting miRNA-mRNA pairs might be missed.

The APA-miRNA study represents an exceptional feature of miRNA binding sites in terms of the time scale. Most features that contribute to a binding site's functionality are absolute during the physiological time scale of an organism, and can be changed mainly during evolution (for example – the sequence of the binding site, or additional sequences around it). Alternative polyadenylation adds a miRNA regulation layer that can be toggled during the physiological time scale of the organism: a binding site that is located prior to the APA site is non-functional, until an APA event occurs, and the site now becomes located near the new end of the 3'UTR, and can be more responsive to the miRNA machinery. Such on-off switch is unique and differs from other features of functional miRNA binding sites.

To conclude, both features – being inside Alus and the position relative to the APA site- can function on-off switches for miRNA functionality. However, the binding sites within Alus can be switched on or off only during evolution, while the binding sites prior to APA sites can be turned on or off during a physiological time scale of a cell or an organism.

How can one continue the work presented in this thesis, and what is the next step in global and local analysis of miRNA-mRNA interactions? Some experiments can be

envisaged to further establish the theories presented in this thesis. For example, a large scale reporter assay of short and long 3'UTRs. If such reporters (of many genes) are cloned, by over-expression or knock-down of miRNAs with binding sites prior to the short 3'UTR end only, the strength of the binding sites can be assessed and compared between the short and long versions. I started doing such experiment with 3 candidate genes; however, many more examples are needed. This experiment can be taken further to check the relative responsiveness of the binding sites under different conditions and cell lines, and try to establish a conclusion about the biological importance of each binding site in the corresponding gene. It will also be of further value to perform global directed experiments, monitoring the functionality of miRNA binding sites and their preferential engagement within the context of the shorter transcript, through the use of methods such as CLIP analysis. Another interesting aspect to continue with is the global shortening. Today there are at the most a few dozens of 3'seq experimental datasets in different cellular systems. However, as time progresses, there might be more and more such datasets, which may be analyzed to find more conditions of global 3'UTR shortening, and in these systems look at the contribution of miRNA binding sites that might become active during APA events. Another type of data is 3'seq analysis of cells undergoing differentiation or induced proliferation. Such data can be examined for global shortening and for the effect of miRNA binding sites. In such experimental systems, a knockdown of important factors for APA (for example the RNA binding protein PABNP1[91]) can introduce more or less shortening, and this can be done together with over-expression or knockdown of important miRNAs. In such design, the levels of the short vs. long 3'UTR of genes with binding sites for the relevant miRNAs can be assessed, to see whether indeed the miRNAs exert stronger inhibitory influence on the short variant.

To continue the interesting search for primate-specific regulation conferred by Alu sequences, one can follow the features of the binding sites within Alus in *Mdm2* and *Mdm4*, especially the multiplicity of binding sites, and look for other such cases. These miRNA-mRNA pairs are of specific interest, in my opinion, since they represent a partial answer to what distinguishes primates and specifically humans from the rest of the animal kingdom. Such experiments can start from a bioinformatics global analysis, to find genes with multiple binding sites for the same miRNA within Alus, and continue to molecular experiments to validate the interactions and find their biological implications. This direction of search has already

been partially started by Spengler et al[160]. This can be taken a step further, to look for genes with multiple binding sites within Alus that match sets of miRNAs that might work together (because of similar expression patterns, for example).

Another important aspect, not researched by me, is the interplay between RNA binding proteins, Alus and miRNA binding sites. Proteins that bind specifically to Alu sequences, in a region overlapping or adjacent to a miRNA target site, might play an important role in the relationship between Alus and miRNAs. They may mask miRNA binding sites within Alus from being functional, and therefore can serve as another mechanism to prevent miRNA binding sites within Alus from being functional. Alternatively, if binding outside the miRNA target site, they may change the structure of the RNA such that the target site either becomes more accessible or less accessible to the miRNA and RNA silencing machinery. However, unlike two of the mechanisms that I identified – the location of the Alu and the structure of the miRNA, such effects can easily be reversed by removal of the binding proteins. Such removal is important experimentally and biologically. Once such a protein is identified, it can be depleted from the system and then a new analysis of miRNA binding site functionality within Alus can be performed (with the limitation that the other masking mechanisms that I described are still present). Such proteins would also be very interesting biologically: under conditions where they are less expressed, binding sites within Alus may suddenly become more accessible to the miRNA machinery, and the contribution of such sites might be greater than previously appreciated. One candidate to start with can be STAU1[161].

In sum, my findings highlight several novel aspects in the regulation and fine tuning of the miRNA-target mRNA interaction, and point out new interesting directions that merit further investigation.

6.1 Perspective: MicroRNAs silence the noisy genome

Yonit Hoffman, Yitzhak Pilpel

Science 03 Apr 2015

All molecular machines have imperfections, and the biological ones are no exception. One type of flaw is a quantitative one: Although all the cells within an organ are genetically identical, the concentrations of many of their proteins can be "noisy" - that is, vary and fluctuate between all the cells. Biologists decompose such noise into two sources: an intrinsic one, which results from the stochastic nature of the biochemistry operating within cells, and an extrinsic one that manifests global differences between cells, such as the number of protein production facilities (e.g., ribosomes)[162]. A major question is whether organisms have evolved means to control noise, especially when imprecisions are detrimental. On page 128 in this issue, Schmiedel *et al.*[163] report combining mathematical modeling and a synthetic gene approach to establish a complex role for microRNAs (miRNAs) in controlling cellular protein content.

Since their discovery, miRNAs have been considered important regulators of basic cellular and organismal biology. These small noncoding RNAs base pair with complementary sequences in messenger RNAs (mRNAs), thereby degrading their mRNA targets or preventing their translation into proteins. Yet, the observation that the quantitative effect of miRNAs on their targets is often minor remains a mystery. It has thus been suggested that miRNAs provide noise filtration functions, limiting variability in protein expression across a population of cells[164,165]. But how can one reveal the potential noise-reducing effect of miRNAs on genes? A mere inspection of genes within their natural complex genomic context might not suffice because this context consists of numerous variables and it is impossible to dissect the effects of each of them. Schmiedel *et al.* avoid these obstacles by analyzing a reporter gene that is synthetically connected to gene parts that convey regulation by miRNA. In particular, the authors constructed a fluorescence reporter that allows measuring of gene expression noise, while varying miRNA regulatory input. In this approach, miRNAs bind to targeted mRNAs through dedicated regions - the 3′-untranslated regions (UTRs) of the mRNAs. Sequences that contain different 3′UTRs, each with one or more binding sites (of varying binding strengths) for different miRNAs, were synthesized. These sequences were each fused to the fluorescent reporter gene. Each

construct was then expressed in cultured mammalian cells (including constructs with no binding site for miRNAs).

Comparing single-cell fluorescence revealed an important difference between reporters that have or that do not have miRNA binding sites. In cells that happened to express the reporter at a low level, noisiness of its expression dropped if the reporter had a miRNA binding site. By contrast, in cells that expressed the reporter at a high level, the presence of a miRNA binding site was associated with elevated noisiness of its expression (see the figure). This result was recapitulated by a mathematical model that implements basic principles of gene expression, with clear predictions: Reduction in intrinsic noise should be proportional to miRNA-mediated repression, and extrinsic noise will be "inherited" from noise in the miRNAs (there is variability in the expression of miRNAs as well). To test the intrinsic noise prediction, Schmiedel *et al.* created another reporter, subject to the same miRNA regulation. Because mRNAs encoding both reporters "see" the same miRNAs, differences between their noise must be ascribed to the intrinsic component. For each reporter, the authors synthesized a version encoding a 3′UTR with or without binding sites for miRNA. The result was clear: miRNA reduced intrinsic noise, even when the reporter was expressed at a high level. This suggests that the original observation - that there is increased noise of a gene's expression when its expression level is high - must have been due to extrinsic noise.

Indeed, as for the extrinsic noise, Schmiedel *et al.* suspected that modifying the noise level of the miRNAs themselves would affect the reporter's noise too. For that, the authors examined what happens if the miRNA is produced from two gene copies, rather than from one. This situation could reduce noise in the miRNA because fluctuations in the expression of one copy are counteracted by the other. They found that miRNAs encoded by more than one gene copy in the genome presented less noise. Further, mRNAs of natural genes are often targeted by more than one type of miRNA. Schmiedel *et al.* determined that such combinatorial effects reduce the amount of the extrinsic noise because it decreases the total amount of miRNA-pool noise. This finding was found to hold also for native genes' 3′UTR.

A key question in any such synthetic approach is, how applicable are the conclusions to natural genes? Examining expression for the entire mouse genome, Schmiedel *et al.* reveal that some 90% of the genes fall within the range of expression that would subject them to such a miRNA-based noise dampening mechanism.

Which genes should be the prime subjects of such a noise dampening mechanism? Single-cell transcriptomics[166,167] should allow noise measurement for each gene and miRNA. With such data, it will be possible to examine the connection between the extent of miRNA regulation of a gene and its noise. Means to manipulate miRNA levels[168] should allow examination of the effect of changes in miRNA expression on the noisiness in their targets. One can then ask which genes are endowed with noise filtration and whether there are genes that are deliberately noisy. Schmiedel *et al.* ascribed intrinsic noise reduction to enhanced transcription that presumably compensates for the mRNA degradation (which maintains a given expression level). Recent reports on the "circular" nature of gene expression - namely, that mRNA degradation feeds back to elevate transcription[169] - may thus provide an intriguing potential mechanism that explains the intrinsic noise reduction effect. And the story need not end with miRNAs. A most profound revolution in genomics is the realization that there are many additional types of RNA. For instance, "antisense" RNAs may also act in noise filtration, especially when coregulated with their corresponding sense transcript[170]. Perhaps some long noncoding RNAs[171], too, contribute to fine tuning of gene expression programs.

7. Acknowledgments

I thank Prof. Tzachi Pilpel and Prof. Moshe Oren for teaching and guiding me how to do great research, computational, statistical and molecular, and how to think outside the box and never give up.

I thank Debora Bublik for investing her time and efforts in teaching me how to do molecular science, for great scientific discussions, good ideas and great friendship.

I thank the following people for their generous help in discussions, ideas, and teaching me new techniques: Dvir Dahary, Yael Aylon, Gilad Fuchs, Hila Gingold, Tammy Biniashvili and Sivan Navon.

I thank Dr. Debora Marks for inspiring discussions and bioinformatics ideas, and for introducing to me the academic world of Harvard and MIT.

I thank Prof. Judy Lieberman and Dr. Paco Navarro for generous help with pull-down expertise.

I thank the Pilpel and Oren labs for discussions and ideas, and for good friendships.

8. Declaration

I declare that the thesis summarizes my independent research.

In the project “The majority of endogenous microRNA binding sites within Alu elements avoid the microRNA machinery” Dvir Dahary contributed in conceiving and designing the research. In the project “miR-661 downregulates both Mdm2 and Mdm4 to activate p53” Debora Bublik contributed in conceiving the research, and performing Cell migration assays.

9. List of Publications

Hoffman Y, Bublik DR, P Ugalde A, Elkon R, Biniashvili T, Agami R, Oren M, Pilpel Y. 3'UTR Shortening Potentiates MicroRNA-Based Repression of Pro-differentiation Genes in Proliferating Human Cells. PLoS Genet. 2016 Feb 23;12(2):e1005879.

Hoffman Y, Pilpel Y. MicroRNAs silence the noisy genome. Science 2015 Apr 3;348(6230):41-2. (Perspective)

Hoffman Y, Bublik DR, Pilpel Y, Oren M. miR-661 downregulates both Mdm2 and Mdm4 to activate p53. Cell Death Differ. 2014;21(2):302-9.

Hoffman Y, Dahary D, Bublik DR, Oren M, Pilpel Y. The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery. Bioinformatics. 2013;29(7):894-902.

10. Bibliography

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

2 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).

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

4 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).

5 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).

6 Hausser, J., Landthaler, M., Jaskiewicz, L., Gaidatzis, D. & Zavolan, M. Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res* **19**, 2009-2020, doi:10.1101/gr.091181.109 (2009).

7 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, doi:10.7554/eLife.05005 (2015).

8 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).

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

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

11 Virant-Klun, I., Stahlberg, A., Kubista, M. & Skutella, T. MicroRNAs: From Female Fertility, Germ Cells, and Stem Cells to Cancer in Humans. *Stem cells international* **2016**, 3984937, doi:10.1155/2016/3984937 (2016).

12 Deng, K., Wang, H., Guo, X. & Xia, J. The cross talk between long, non-coding RNAs and microRNAs in gastric cancer. *Acta biochimica et biophysica Sinica*, doi:10.1093/abbs/gmv120 (2015).

13 Cekaite, L., Eide, P. W., Lind, G. E., Skotheim, R. I. & Lothe, R. A. MicroRNAs as growth regulators, their function and biomarker status in colorectal cancer. *Oncotarget*, doi:10.18632/oncotarget.6390 (2015).

14 Ameres, S. L. & Zamore, P. D. Diversifying microRNA sequence and function. *Nature reviews. Molecular cell biology* **14**, 475-488, doi:10.1038/nrm3611 (2013).

15 Kobayashi, H. & Tomari, Y. RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochimica et biophysica acta*, doi:10.1016/j.bbagrm.2015.08.007 (2015).

16 Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642-655, doi:10.1016/j.cell.2009.01.035 (2009).

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

18 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).
19 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703 (2009).
20 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).
21 Deininger, P. Alu elements: know the SINEs. *Genome Biol* **12**, 236, doi:10.1186/gb-2011-12-12-236 (2011).
22 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).
23 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).
24 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).
25 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).
26 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).
27 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).
28 Lev-Maor, G. *et al.* Intronic Alus influence alternative splicing. *PLoS Genet* **4**, e1000204, doi:10.1371/journal.pgen.1000204 (2008).
29 Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. *Genome Res* **12**, 1060-1067, doi:10.1101/gr.229302 (2002).
30 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).
31 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).
32 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).
33 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).

34 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).
35 Piriyapongsa, J., Marino-Ramirez, L. & Jordan, I. K. Origin and evolution of human microRNAs from transposable elements. *Genetics* **176**, 1323-1337 (2007).
36 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).
37 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).
38 Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**, 1097-1101 (2006).
39 Bortolin-Cavaille, M. L., Dance, M., Weber, M. & Cavaille, J. C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. *Nucleic Acids Res* **37**, 3464-3473, doi:gkp205 [pii] 10.1093/nar/gkp205 (2009).
40 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).
41 Smalheiser, N. R. & Torvik, V. I. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet* **22**, 532-536 (2006).
42 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).
43 Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nat Rev Genet* **3**, 370-379 (2002).
44 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).
45 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).
46 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).
47 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).
48 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).
49 Weinberg, W. C. & Denning, M. F. P21Waf1 control of epithelial cell cycle and cell fate. *Crit Rev Oral Biol Med* **13**, 453-464 (2002).
50 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).

51 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).
52 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).
53 Ferbeyre, G. *et al.* PML is induced by oncogenic ras and promotes premature senescence. *Genes Dev* **14**, 2015-2027 (2000).
54 Pearson, M. *et al.* PML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature* **406**, 207-210, doi:10.1038/35018127 (2000).
55 Tan, T. & Chu, G. p53 Binds and activates the xeroderma pigmentosum DDB2 gene in humans but not mice. *Mol Cell Biol* **22**, 3247-3254 (2002).
56 Hwang, B. J., Ford, J. M., Hanawalt, P. C. & Chu, G. Expression of the p48 xeroderma pigmentosum gene is p53-dependent and is involved in global genomic repair. *Proc Natl Acad Sci U S A* **96**, 424-428 (1999).
57 Christmann, M. & Kaina, B. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. *Nucleic Acids Res* **41**, 8403-8420, doi:10.1093/nar/gkt635 (2013).
58 Dashzeveg, N. & Yoshida, K. Cell death decision by p53 via control of the mitochondrial membrane. *Cancer Lett* **367**, 108-112, doi:10.1016/j.canlet.2015.07.019 (2015).
59 Wang, X., Simpson, E. R. & Brown, K. A. p53: Protection against Tumor Growth beyond Effects on Cell Cycle and Apoptosis. *Cancer Res* **75**, 5001-5007, doi:10.1158/0008-5472.CAN-15-0563 (2015).
60 Kruse, J. P. & Gu, W. Modes of p53 regulation. *Cell* **137**, 609-622, doi:10.1016/j.cell.2009.04.050 (2009).
61 Wade, M., Li, Y. C. & Wahl, G. M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat Rev Cancer* **13**, 83-96, doi:nrc3430 [pii] 10.1038/nrc3430 (2013).
62 Hu, W., Feng, Z. & Levine, A. J. The Regulation of Multiple p53 Stress Responses is Mediated through MDM2. *Genes Cancer* **3**, 199-208, doi:10.1177/1947601912454734 10.1177_1947601912454734 [pii] (2012).
63 Li, Q. & Lozano, G. Molecular pathways: targeting Mdm2 and Mdm4 in cancer therapy. *Clin Cancer Res* **19**, 34-41, doi:1078-0432.CCR-12-0053 [pii] 10.1158/1078-0432.CCR-12-0053 (2013).
64 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).
65 Urso, L., Calabrese, F., Favaretto, A., Conte, P. & Pasello, G. Critical review about MDM2 in cancer: Possible role in malignant mesothelioma and implications for treatment. *Crit Rev Oncol Hematol*, doi:10.1016/j.critrevonc.2015.08.019 (2015).
66 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).
67 Barak, Y., Juven, T., Haffner, R. & Oren, M. mdm2 expression is induced by wild type p53 activity. *EMBO J* **12**, 461-468 (1993).

68 Wu, X., Bayle, J. H., Olson, D. & Levine, A. J. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev* **7**, 1126-1132 (1993).
69 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).
70 Shvarts, A. *et al.* MDMX: a novel p53-binding protein with some functional properties of MDM2. *EMBO J* **15**, 5349-5357 (1996).
71 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).
72 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).
73 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).
74 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).
75 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).
76 Kim, E. S. & Shohet, J. M. Reactivation of p53 via MDM2 inhibition. *Cell death & disease* **6**, e1936, doi:10.1038/cddis.2015.302 (2015).
77 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).
78 Li, Z., Liu, Y. H., Diao, H. Y., Ma, J. & Yao, Y. L. MiR-661 inhibits glioma cell proliferation, migration and invasion by targeting hTERT. *Biochem Biophys Res Commun* **468**, 870-876, doi:10.1016/j.bbrc.2015.11.046 (2015).
79 Zhu, T. *et al.* MiR-661 contributed to cell proliferation of human ovarian cancer cells by repressing INPP5J expression. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **75**, 123-128, doi:10.1016/j.biopha.2015.07.023 (2015).
80 Rutnam, Z. J., Du, W. W., Yang, W., Yang, X. & Yang, B. B. The pseudogene TUSC2P promotes TUSC2 function by binding multiple microRNAs. *Nature communications* **5**, 2914, doi:10.1038/ncomms3914 (2014).
81 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).
82 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).
83 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).
84 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).
85 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).
86 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).
87 Elkon, R., Ugalde, A. P. & Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**, 496-506, doi:10.1038/nrg3482 (2013).
88 Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**, 2757-2774, doi:10.1093/nar/gkp1176 (2010).
89 Shi, Y. Alternative polyadenylation: new insights from global analyses. *RNA* **18**, 2105-2117, doi:10.1261/rna.035899.112 (2012).
90 Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A. & Moreira, A. Implications of polyadenylation in health and disease. *Nucleus* **5**, 508-519, doi:10.4161/nucl.36360 (2014).
91 Jenal, M. *et al.* The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**, 538-553, doi:10.1016/j.cell.2012.03.022 (2012).
92 Elkon, R. *et al.* E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol* **13**, R59, doi:10.1186/gb-2012-13-7-r59 (2012).
93 Mayr, C. & Bartel, D. P. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673-684, doi:10.1016/j.cell.2009.06.016 (2009).
94 Lin, Y. *et al.* An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res* **40**, 8460-8471, doi:10.1093/nar/gks637 (2012).
95 Lembo, A., Di Cunto, F. & Provero, P. Shortening of 3'UTRs correlates with poor prognosis in breast and lung cancer. *PLoS One* **7**, e31129, doi:10.1371/journal.pone.0031129 (2012).
96 Ji, Z. & Tian, B. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* **4**, e8419, doi:10.1371/journal.pone.0008419 (2009).
97 Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol* **6**, R100, doi:10.1186/gb-2005-6-12-r100 (2005).
98 Mayr, C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol*, doi:10.1016/j.tcb.2015.10.012 (2015).

99 To, K. K. *et al.* Escape from hsa-miR-519c enables drug-resistant cells to maintain high expression of ABCG2. *Molecular cancer therapeutics* **8**, 2959-2968, doi:10.1158/1535-7163.MCT-09-0292 (2009).
100 Tranter, M. *et al.* Coordinated post-transcriptional regulation of Hsp70.3 gene expression by microRNA and alternative polyadenylation. *J Biol Chem* **286**, 29828-29837, doi:10.1074/jbc.M111.221796 (2011).
101 Neve, J. & Furger, A. Alternative polyadenylation: less than meets the eye? *Biochemical Society transactions* **42**, 1190-1195, doi:10.1042/BST20140054 (2014).
102 Nam, J. W. *et al.* Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol Cell* **53**, 1031-1043, doi:10.1016/j.molcel.2014.02.013 (2014).
103 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).
104 Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144, doi:10.1093/nar/gkj112 (2006).
105 Khan, A. A. *et al.* Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* **27**, 549-555 (2009).
106 Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911-940, doi:10.1006/jmbi.1999.2700 (1999).
107 Wuchty, S., Fontana, W., Hofacker, I. L. & Schuster, P. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**, 145-165, doi:10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G (1999).
108 Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature methods* **8**, 559-564, doi:10.1038/nmeth.1608 (2011).
109 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
110 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).
111 Nielsen, C. B. *et al.* Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894-1910, doi:10.1261/rna.768207 (2007).
112 Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol* **3**, e85 (2005).
113 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).
114 Ender, C. & Meister, G. Argonaute proteins at a glance. *J Cell Sci* **123**, 1819-1823, doi:123/11/1819 [pii] 10.1242/jcs.055210 (2010).

115 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).
116 Blow, M., Futreal, P. A., Wooster, R. & Stratton, M. R. A survey of RNA editing in human brain. *Genome Res* **14**, 2379-2387 (2004).
117 Kim, D. D. *et al.* Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res* **14**, 1719-1725 (2004).
118 Levanon, E. Y. *et al.* Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**, 1001-1005 (2004).
119 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).
120 Barak, M. *et al.* Evidence for large diversity in the human transcriptome created by Alu RNA editing. *Nucleic Acids Res* **37**, 6905-6915 (2009).
121 Borchert, G. M. *et al.* Adenosine deamination in human transcripts generates novel microRNA binding sites. *Hum Mol Genet* **18**, 4801-4807, doi:ddp443 [pii]
10.1093/hmg/ddp443 (2009).
122 Ullu, E. & Tschudi, C. Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171-172 (1984).
123 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).
124 Aylon, Y. *et al.* The Lats2 tumor suppressor augments p53-mediated apoptosis by promoting the nuclear proapoptotic function of ASPP1. *Genes Dev* **24**, 2420-2429 (2010).
125 Golomb, L. *et al.* Importin 7 and exportin 1 link c-Myc and p53 to regulation of ribosomal biogenesis. *Mol Cell* **45**, 222-232 (2012).
126 Orom, U. A. & Lund, A. H. Isolation of microRNA targets using biotinylated synthetic microRNAs. *Methods* **43**, 162-165 (2007).
127 Lal, A. *et al.* Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet* **7**, e1002363 (2011).
128 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:2/5/401 [pii]
10.1158/2159-8290.CD-12-0095 (2012).
129 Antonov, A. V., Knight, R. A., Melino, G., Barlev, N. A. & Tsvetkov, P. O. MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. *Cell Death Differ* **20**, 367 (2012).
130 Lyng, M. B. *et al.* Global microRNA expression profiling of high-risk ER+ breast cancers from patients receiving adjuvant tamoxifen mono-therapy: a DBCG study. *PLoS One* **7**, e36170, doi:10.1371/journal.pone.0036170 (2012).
131 Hoffman, Y., Dahary, D., Bublik, D. R., Oren, M. & Pilpel, Y. The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery. *Bioinformatics* **29**, 894-902 (2013).
132 O'Connor, P. M. *et al.* Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and

correlations with the growth-inhibitory potency of 123 anticancer agents. *Cancer Res* **57**, 4285-4300 (1997).
133 Antonov, A. V., Knight, R. A., Melino, G., Barlev, N. A. & Tsvetkov, P. O. MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. *Cell Death Differ* **20**, 367, doi:10.1038/cdd.2012.137 (2013).
134 Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-10874 (2001).
135 Ahmed, A. A. *et al.* Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J Pathol* **221**, 49-56 (2010).
136 Bernardini, M. Q. *et al.* Expression signatures of TP53 mutations in serous ovarian cancers. *BMC Cancer* **10**, 237, doi:1471-2407-10-237 [pii] 10.1186/1471-2407-10-237 (2010).
137 Oren, M. & Rotter, V. Mutant p53 gain-of-function in cancer. *Cold Spring Harb Perspect Biol* **2**, a001107, doi:10.1101/cshperspect.a001107 (2010).
138 Walerych, D., Napoli, M., Collavin, L. & Del Sal, G. The rebel angel: mutant p53 as the driving oncogene in breast cancer. *Carcinogenesis* **33**, 2007-2017, doi:bgs232 [pii] 10.1093/carcin/bgs232 (2012).
139 Muller, P. A., Vousden, K. H. & Norman, J. C. p53 and its mutants in tumor cell migration and invasion. *J Cell Biol* **192**, 209-218 (2011).
140 Lopez-Romero, P. Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library. *BMC Genomics* **12**, 64, doi:10.1186/1471-2164-12-64 (2011).
141 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
142 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121, doi:10.1101/gr.097857.109 (2010).
143 Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-1050, doi:10.1101/gr.3715005 (2005).
144 Gingold, H. *et al.* A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**, 1281-1292, doi:10.1016/j.cell.2014.08.011 (2014).
145 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* **6**, 26, doi:10.1186/1748-7188-6-26 (2011).
146 Milyavsky, M. *et al.* Prolonged culture of telomerase-immortalized human fibroblasts leads to a premalignant phenotype. *Cancer Res* **63**, 7147-7157 (2003).
147 Aylon, Y. *et al.* Silencing of the Lats2 tumor suppressor overrides a p53-dependent oncogenic stress checkpoint and enables mutant H-Ras-driven cell transformation. *Oncogene* **28**, 4469-4479, doi:10.1038/onc.2009.270 (2009).
148 Elkon, R. *et al.* E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol* **13**, R59, doi:10.1186/gb-2012-13-7-r59 (2012).
149 Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105, doi:10.1101/gr.082701.108 (2009).

150 Mogilyansky, E. & Rigoutsos, I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* **20**, 1603-1614, doi:10.1038/cdd.2013.125 (2013).
151 Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696-700, doi:10.1038/nature12756 (2014).
152 Sood, P., Krek, A., Zavolan, M., Macino, G. & Rajewsky, N. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci U S A* **103**, 2746-2751, doi:10.1073/pnas.0511045103 (2006).
153 Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654-665, doi:10.1016/j.cell.2013.03.043 (2013).
154 Zhang, R., Peng, Y., Wang, W. & Su, B. Rapid evolution of an X-linked microRNA cluster in primates. *Genome Res* **17**, 612-617 (2007).
155 Flor, I. & Bullerdiek, J. The dark side of a success story: microRNAs of the C19MC cluster in human tumours. *J Pathol* **227**, 270-274 (2012).
156 Shalgi, R., Pilpel, Y. & Oren, M. Repression of transposable-elements - a microRNA anti-cancer defense mechanism? *Trends Genet* **26**, 253-259 (2010).
157 Ding, J., Li, D., Ohler, U., Guan, J. & Zhou, S. Genome-wide search for miRNA-target interactions in Arabidopsis thaliana with an integrated approach. *BMC Genomics* **13 Suppl 3**, S3, doi:10.1186/1471-2164-13-S3-S3 (2012).
158 Appasani, K. MicroRNAs FROM BASIC SCIENCE TO DISEASE BIOLOGY Introduction. *Micrornas: From Basic Science to Disease Biology*, 1-3 (2008).
159 Le Bras, M., Bensaad, K. & Soussi, T. Data mining the p53 pathway in the Fugu genome: evidence for strong conservation of the apoptotic pathway. *Oncogene* **22**, 5082-5090, doi:10.1038/sj.onc.1206424 (2003).
160 Spengler, R. M., Oakley, C. K. & Davidson, B. L. Functional microRNAs and target sites are created by lineage-specific transposition. *Hum Mol Genet* **23**, 1783-1793, doi:10.1093/hmg/ddt569 (2014).
161 Elbarbary, R. A., Li, W., Tian, B. & Maquat, L. E. STAU1 binding 3' UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev* **27**, 1495-1510, doi:10.1101/gad.220962.113 (2013).
162 Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186, doi:10.1126/science.1070919 (2002).
163 Schmiedel, J. M. *et al.* Gene expression. MicroRNA control of protein expression noise. *Science* **348**, 128-132, doi:10.1126/science.aaa1738 (2015).
164 Hornstein, E. & Shomron, N. Canalization of development by microRNAs. *Nat Genet* **38 Suppl**, S20-24, doi:10.1038/ng1803 (2006).
165 Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149**, 515-524, doi:10.1016/j.cell.2012.04.005 (2012).
166 Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776-779, doi:10.1126/science.1247651 (2014).

167 Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports* **2**, 666-673, doi:10.1016/j.celrep.2012.08.003 (2012).
168 Krutzfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685-689, doi:10.1038/nature04303 (2005).
169 Haimovich, G. *et al.* Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* **153**, 1000-1011, doi:10.1016/j.cell.2013.05.012 (2013).
170 Lapidot, M. & Pilpel, Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* **7**, 1216-1222, doi:10.1038/sj.embor.7400857 (2006).
171 Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227, doi:10.1038/nature07672 (2009).