מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE

עבודת גמר (תזה) לתואר

דוקטור לפילוסופיה

מוגשת למועצה המדעית של

מכון ויצמן למדע

רחובות, ישראל

Thesis for the degree

Doctor of Philosophy

Submitted to the Scientific Council of the

Weizmann Institute of Science

Rehovot, Israel

במתכונת "מאמרים"

In a "Published Papers" Format

מאת

רעות שלגי

By

Reut Shalgi

הקשר בין בקרה ברמת השיעתוק לבין בקרה על ידי מיקרו-רנ"א, ותפקידם בויסות פרוליפרציה של התא

Coupling transcriptional and post-transcriptional miRNA regulation in the control of cell proliferation

מנחים:

פרופ' משה אורן

פרופ' יצחק פלפל

Advisors:

Prof. Moshe Oren

Prof. Yitzhak Pilpel

אלול תשס"ט

September 2009

Table of Contents

Abstract

In recent years it became evident that miRNA regulation plays a role in numerous developmental and cellular processes. Similarly to other well known regulatory levels, post-transcriptional regulation by miRNAs appears as an integral part of the regulatory network of virtually almost all biological processes in eukaryotic organisms. Aiming to explore the structure of the regulatory networks in which miRNAs are embedded, I investigated the possibility of global coordination between transcription and post-transcription regulation by miRNAs.

I found two major principles of miRNA regulatory network structure: The first is miRNA combinatorial regulation, i.e. the functional regulatory interaction between miRNAs, which I demonstrated computationally, and which was supported by other computational and experimental studies henceforth. The second is the coupling of transcription regulation with post-transcriptional regulation by miRNAs as a recurrent network architecture, showing that often in the network miRNAs and transcription factors (TFs) significantly regulate joint targets, while also regulating each other in the form of Feed Forward Loops (FFLs). This study was accompanied by predictions for specific co-regulating pairs and FFLs. Focusing on the latter I further explored the involvement of such FFLs in the regulation of cell fate, specifically proliferation and senescence. We investigated a particular case of a FLL prediction consisting of the cell cycle regulator *E2F* and a family of 15 miRNAs, validated the prediction experimentally, and showed that it is involved in the regulation of cell proliferation. The miRNAs in the family combinatorially regulate a set of *E2F* target genes, many of which are regulators of cell cycle. We revealed a link to cancer, as the regulators of this FFL are repressed by *p53* when normal cells enter replicative senescence, and when *p53* is mutated in breast cancer, this miRNA family is upregulated. Perturbations in the FFL, in the form of miRNA over-expression, caused enhanced proliferation and delayed senescence, attesting to its crucial role in the maintenance of proper proliferation in normal cells. I continued to explore the role of TF-miRNA coupling in the regulation of cell fate, demonstrating its role in oncogene-induced senescence (Christoffersen *et al.* 2009), and in *EGF* signaling

pathways (Avraham *et al.* 2009) in two collaborative studies. During this time publications have appeared on other combined TF-miRNA FFLs and their involvement in cell fate in various organisms and cell systems (reviewed in (Shalgi et al., 2009)).

Another direction of my research explores the interesting question of the evolutionary origin of animal miRNAs. In this study, two alternative genomic sources which serve as a constant supply for novel miRNAs during evolution are described. The first is Transposable elements (TEs), which were previously described in the literature as a class of genomic elements serving as an origin for miRNA innovations. The second source is a newly described possible origin for miRNAs in evolution: CpG islands (CGIs) (Dahary*, Shalgi* et al., 2009).

Finally, integrating both avenues of my work in the context of miRNAs and growth control, with the new line of investigation of miRNA evolution, I recently suggested possible new roles for miRNAs in the maintenance of genomic integrity and cell fate in normal cells. I suggest that the coupling of miRNAs with transcription regulation in the network, is a crucial aspect in the significant effect miRNAs have on cell fate. In addition, I hypothesize that TE-originated miRNAs might be directly involved in maintaining genomic integrity via global repression of TEs. While the coupling of transcription with miRNA regulation as a principle of mammalian regulatory networks notion was well substantiated during my PhD work, both by myself and by others, the latter notion still remains to be verified. Together my work demonstrated both general principles in miRNA regulatory networks, i.e. miRNA combinatorial regulation and the coupling of transcription and miRNA regulation as recurrent architectures in mammalian regulatory network, and how these principles are manifested in human cells, as part of the complex network that controls cell proliferation and senescence. My work contributed towards the understanding of miRNA regulatory networks, and how those significantly affect cell fate.

סיכום

בשנים האחרונות התברר שבקרה על ידי מיקרו-רנ"א משחקת תפקיד מכריע בתהליכים התפתחותיים ותאיים רבים. בדומה לשכבות בקרה אחרות, נראה שבקרה פוסט-שיעתוק ע"י מיקרו-רנ"א הינה חלק אינטגרלי מרשתות הבקרה של כמעט כל התהליכים הביולוגים ביצורים אאוקריוטיים. במטרה לחקור את מבנה רשתות הבקרה הכוללות מיקרו-רנ"א, חקרתי את הקשר שבין בקרת שיעתוק לבין בקרת פוסט-שיעתור על ידי מיקרו-רנ"א.

גיליתי שני עקרונות מרכזיים של מבנה רשת הבקרה ע"י מיקרו-רנ"א. הראשון הוא בקרה קומבינטורית, כלומר אינטראקציה פונקציונלית בין מיקרו-רנ"א, אותו הדגמתי חישובית, ונתמך ע"י מחקרים חישוביים וניסיוניים נוספים. השני הינו הקשר בין בקרת שיעתוק לבין בקרת פוסט-שיעתוק ע"י מיקרו-רנ"א, כארכיטקטורה חוזרנית ברשת, כאשר פקטורי שיעתוק ומיקרו-רנ"א נוטים לבקר גני-מטרה משותפים, ובנוסף לבקר זה את זה בצורה של מעגל היזון-קדימה (feed forward loop, FFL). למחקר זה התלוותה רשימה של תחזיות של זוגות פקטורי שיעתוק ומיקרו-רנ"א היוצרים בקרה משותפת וFFLs.

התמקדתי בעיקרון השני, והמשכתי להתעמק ולחקור את המעורבות של מעגלים מסוג זה בבקרה של גורל התא, ובאופן ספיציפי בפרוליפרציה והזדקנות תאית. במסגרת זו חקרתי מקרה פרטי של מעגל חזוי שכזה, המערב את בקר מחזור התא E2F ומשפחה של 15 מיקרו-רנ"א. אימתנו ניסיונית שאכן מעגל זה מתקיים בתא, והדגמתי את מעורבותו בבקרה של פרוליפרציה תאית. משפחת המיקרו-רנ"א המדוברת מבקרת קומבינטורית קבוצה של גנים הידועים כמטרות בקרתיות של E2F ברמת השיעתוק, ורבים מהם הינם בעצמם בקרים של מחזור התא. בנוסף גילינו קשר לסרטן: הבקרים של מעגל זה מדוכאים על ידי p53 בעת שתאים נורמאליים נכנסים להזדקנות חלוקתית, וכן נצפתה עליה ברמת המיקרו-רנ"א המדוברים בדגימות סרטן שד בהם יש מוטציות ב-p53. פרטורבציות במעגל זה, שהתבצעו ע"י ביטוי ייתר של המיקרו-רנ"א, גרמו לפרוליפרציה מוגברת, ודחייה בהזדקנות התאית. דבר זה מעיד על תפקידו של מעגל זה בשמירה על פרוליפרציה בתאים נורמאליים.

המשכתי לחקור את התפקיד של הקשר בין פקטורי תיעתוק למיקרו-רנ"א בבקרה על גורל התא, במסגרת שני שיתופי פעולה נוספים, והדגמתי את הקשר שלו להזדקנות תאית המושרית ע"י אונקוגן (oncogene-induced senescence), ובנוסף במסלולי בקרה ע"י EGF. במהלך התקופה, פרסומים נוספים הדגימו מגוון של FFLs של פקטורי תיעתוק ומיקרו-רנ"א ומעורבותם בבקרה על גורל התא במגוון אורגניזמים ומערכות תאיות.

כיוון נוסף אותו חקרתי הינו השאלה המעניינת של "איך מיקרו-רנ"א נולד?" – כלומר המקור האבולוציוני של מיקרו-רנ"א. במחקר זה, הצגנו שתי אלטרנטיבות למקור גנומי שהיוו אספקה של מיקרו-רנ"א חדשים במהלך האבולוציה. הראשון הוא טרנספוזונים, שתוארו בעבר בספרות כקבוצה של אלמנטים גנומיים ששימשו כמקור ליצירת מיקרו-רנ"א חדשים באבולוציה. השני הינו מקור המתואר על ידינו לראשונה כמקור פוטנציאלי ליצירת מיקרו-רנ"א חדשים באבולוציה: איי CpG.

לבסוף, חיברתי את שני תחומי העבודה שלי: מיקרו-רנ"א בבקרה על פרוליפרציה תאית מצד אחד, ומצד שני את קו המחקר החדש על המקור האבולוציוני של מיקרו-רנ"א, ולאחרונה הצעתי את הרעיון שקיים תפקיד אפשרי נוסף למיקרו-רנ"א בשמירה על שלמות הגנום ותקינות גורל התא בתאים נורמאליים. הצעתי שהצימוד בין מיקרו-רנ"א לבקרת שיעתוק ברשת היא היבט מכריע בהשפעה המשמעותית שיש למיקרו-רנ"א על גורל התא. בנוסף, העליתי את ההיפותיזה שייתכן שמיקרו-רנ"א שמקורם בטרנספוזונים מעורבים ישירות בשמירה על תקינות הגנום ע"י רפרסיה גלובלית של טרנספוזונים. בעוד שהצימוד בין בקרת שיעתוק עם בקרה ע"י מיקרו-רנ"א כעיקרון משמעותי ברשתות רגולציה ביונקים הינו רעיון שבוסס היטב במהלך עבודת הדוקטורט שלי, באמצעות העבודה שלי וגם באמצעות עבודות של אחרים, את הרעיון האחרון יש עדיין להוכיח.

לסיכום, עבודתי הראתה מחד עקרונות כלליים ברשתות בקרה של מיקרו-רנ"א, כלומר בקרה קומבינטורית ע"י מיקרו-רנ"א וצימוד בין בקרת שיעתוק לבקרה ע"י מיקרו-רנ"א כארכיטקטורה חוזרנית ברשתות בקרה ביונקים, ומאידך כיצד עקרונות אלו באים לידי ביטוי בתאי אדם, כחלק מרשת הבקרה המורכבת של פרוליפרציה והזדקנות. עבודתי תרמה להבנת רשתות בקרה של מיקרו-רנ"א, וכן להבנה כיצד רשתות אילו משפיעות על גורל התא.

Introduction

Introduction to miRNAs

miRNAs are small regulatory RNAs that exert gene silencing in many organisms in, including both plants and animals (Bartel 2004).

Post-transcriptional silencing by miRNAs is guided by the recognition of a complementary site on the target mRNA, followed by either cleavage of the target or its translation inhibition (Bartel 2004). miRNAs exercise post-transcriptional silencing by leading a protein complex named RNA-Induced Silencing Complex (RISC) to cleavage of their target mRNAs and to inhibition of their translation. While in plants the sites may lie all across the transcript, in metazoans the miRNA were mainly believed to target the 3' UTR of the mRNA (Bartel 2004). Yet, a recent study by Chi et al. reported the immunoprecipitation of mouse RISC complexes and the successful sequencing and analysis of both miRNAs and bound target mRNAs, and revealed that, contrary to the current convention, many of the sites bound by RISCs are in the coding region and other regions of the mRNA (Chi *et al.* 2009).

The most important component of the RISC is the *Argonaute* protein (termed *Ago* for short) (Hammond *et al.* 2001). There are 4 different *Argonaute* proteins in the human genome (Sasaki *et al.* 2003). *Argonaute* 's contain a PAZ domain, which is responsible for the binding of the miRNA molecule (Song *et al.* 2003), and a PIWI domain, which is an RNase H endonuclease-like domain. While *Ago1, 2* and *3* were shown to reside within an active RISC, only *Ago2* was found to be capable of target cleavage (Liu *et al.* 2004; Meister *et al.* 2004), and had an active RNase H-like domain. When certain residues in the active site were mutated, *Ago2* lost its cleavage activity (Liu *et al.* 2004). Besides the protein, RISC consists of additional components and interaction partners (Filipowicz 2005; van den Berg *et al.* 2008) (Rana 2007), and part of its silencing activity is

attributed to its recruitment of, or to, P-bodies and P-body components (Parker and Sheth 2007).

miRNA-loaded RISC can exert silencing via mechanism of the target transcript degradation and inhibition of its translation. Mechanistically, it became evident that mammalian *Ago2* is capable of both mRNA cleavage and inhibition of translation (Liu *et al.* 2004; Meister *et al.* 2004), while the other three *Argonaute* family proteins exert silencing only through inhibition of translation. An old convention in the field was that animal miRNAs exert most of their silencing through the inhibition of translation, rather than through the degradation of their targets, as opposed to plant miRNAs in which silencing is mainly mediated by cleavage. This was attributed to a poor overall degree of sequence complementarity that animal miRNAs often share with their target sites on 3' UTRs of mRNAs (Bartel 2004). The first discovered miRNAs in C.elegans, *lin-4*, was shown to inhibit the translation of its target *Lin-14*, without affecting its mRNA levels (Wightman *et al.* 1993; Olsen and Ambros 1999). It is now further established that miRNAs are incorporated into RISCs without a preference towards any of the four Ago proteins (Liu *et al.* 2004). It appeared however that the picture is more complicated. Subsequently, a study by Lim et al. was the first to show that miRNAs can influence the mRNA levels of their target genes. Using overexpression of miRNAs followed by global expression profiling using microarrays, Lim et al. showed a modest but significant downregulation of mRNA levels of genes which were enriched for the miRNA seed sequence (Lim *et al.* 2005). Recent studies used high throughput proteomics in order to both identify translationally inhibited targets and to more accurately assess the extent of inhibition that a miRNA exerts on mRNA levels and on protein levels together (Baek *et al.* 2008; Selbach *et al.* 2008). These studies reported that individual miRNAs affect hundreds of proteins; however the levels of these proteins were decreased only to a relatively mild extent, and short term effects were shown to be mediated mainly at the protein level, whereas long term effects were shown to be both at the mRNA and protein levels, and their magnitude was highly correlated. These studies and others contributed to the overall current view that animal miRNAs exert silencing through both mechanisms simultaneously (Rana 2007).

miRNAs are transcribed from the genome as primary transcripts (termed pri-miRNAs), which can be quite long, up do dozens of kilobases, and are in many cases polycistronic, containing several miRNAs together. miRNA genes reside either in independent loci on the genome, or in introns of known protein coding genes, and much less frequently they may also overlap exons of coding genes. More than a third of all documented human miRNAs are intronic (see Figure 1). Pri-miRNAs were generally shown to be transcribed by RNA polymerase II (Lee *et al.* 2004), and undergo capping and polyadenylation (reviewed in (Kim 2005)). However, there is evidence for transcription of miRNAs by RNA polymerase III as well (Borchert *et al.* 2006; Dieci *et al.* 2007). Pri-miRNAs are processed in the nucleus by the *Drosha-DGCR8* complex (Lee *et al.* 2003) (Han *et al.* 2004), and cleaved to their pre-miRNA form (a hairpin of about 70 bases), which is exported from the nucleus and further cleaved by *Dicer* to its mature, single stranded ~22 bases long form (Cullen 2004) (Kim 2005). All of these steps in the miRNA biogenesis pathway are regulated in various ways, which are only now starting to unravel (reviewed in (Winter *et al.* 2009).

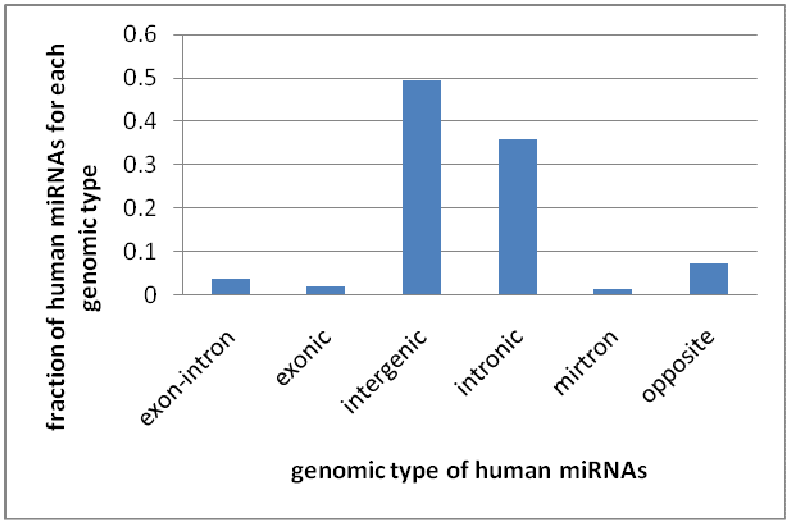


Figure 1

Distribution of types of genomic location of miRNAs in the human genome.

miRNA regulatory networks

Whereas miRNAs, being *PolII* transcribed genes, have promoters that are similar to those of most protein coding genes and are regulated by same transcription factors (TFs), identification of miRNA promoter regions is not a trivial task. pri-miRNAs can be long, in some cases dozens of kilobases (Raver-Shapira *et al.* 2007), and very hard to identify. General bioinformatic based methods have estimated the promoters to be on average ~4kb upstream to the pre-miRNA sequence (Saini *et al.* 2007). However, a recent study that used global ChIP-seq of various histone modifications, mapped many transcription start sites (TSS) of mouse and human pri-miRNAs, and found that only about half of them (in both species) actually have a promoter that is less than 5kb away from the pre-miRNA (Marson *et al.* 2008). Nevertheless, in the past few years many studies were published that identified promoter regions for individual miRNAs, such as in the case of the *mir-23a-24–2* polycistron (Lee *et al.* 2004).Transcriptional regulation of specific miRNA genes by different TFs is a subject of intensive research, and many studies have shown cases of regulation on a one TF-one miRNA basis (O'Donnell *et al.* 2005; He *et al.* 2007; Raver-Shapira *et al.* 2007; Sylvestre *et al.* 2007).

Animal miRNA target identification is also not a trivial task. While in plants, miRNAs usually share near perfect complementarity with their targets, making it easy to identify the targets of each miRNA (Lim *et al.* 2003), animal miRNAs rules of targeting complexity is far from being resolved (reviewed in (Bartel 2009)). Numerous prediction algorithms were developed for miRNA target identification in animals, many of which use evolutionary conservation as a filter for target prediction. The most well-known algorithms are PicTar (Krek *et al.* 2005), miRanda (Enright *et al.* 2003; John *et al.* 2004) and TargetScan (Lewis *et al.* 2003; Lewis *et al.* 2005), but they all still have a fairly high rate of false positive predictions. Several determinants of miRNA targeting were revealed over the years, such as the seed match (denotes the perfect match of nucleotides 2-8 in the 5’ of the miRNA to its target) (Lewis *et al.* 2003), which is very common and can be sufficient for targeting in some cases, but is not necessary in others (Brennecke *et al.*

2005), as well as other determinants (Lewis *et al.* 2005; Grimson *et al.* 2007; Nielsen *et al.* 2007). Experimental approaches for large-scale identification of miRNA targets were also published over the years, such as miRNA over-expression followed by expression microarrays, first presented by Lim et al. (Lim *et al.* 2005). However, this method relied on the miRNA effect on degradation of its targets, which in many cases is extremely mild. Unfortunately, these methods have not matured to a comprehensive systematic effort, and most of the experimentally verified miRNA targets were collected from studies reporting a single target for a single miRNA (Sethupathy *et al.* 2006).

Very recently, a breakthrough in the field of target identification seems to have happened, when two novel large-scale experimental approaches were successfully taken to identify miRNA targets. The first used proteomics methodologies to identify changes in protein levels following the over-expression, knock-down or knock-out of specific miRNAs (Baek *et al.* 2008; Selbach *et al.* 2008). The second has succeeded in immunoprecipitating RISCs and sequencing the mRNAs that were bound to them in a high degree of specificity (Chi *et al.* 2009). It seems that these methods have opened the opportunity for a more specific and accurate identification of miRNA targets, and will be taken systematically for many miRNAs in many biological systems. These would reveal in an accurate manner the missing parts of the miRNA regulation network components, perhaps allowing the field to move forward in miRNA research and better study it in a systems view.

Despite uncertainties in target identification, a growing interest has begun in the last few years in exploring miRNA regulation in the context of regulatory networks. Examining the global network of miRNAs and their predicted targets has provided some very important notions of our understanding of miRNA regulation, such as the “miR-target avoidance” phenomenon (Farh *et al.* 2005; Stark *et al.* 2005). This phenomenon was described based on miRNA-target network identification combined with expression data during embryonic development and adult tissues in fly and mouse, and indicated that miRNAs targets tend to avoid being expressed in the same tissues or the same developmental times as their regulating miRNAs, thus facilitating the robustness of

developmental boundaries and programs (Farh *et al.* 2005; Stark *et al.* 2005) Another important observation made by looking at miRNA-target networks was that miRNAs tend to heavily target transcription regulators (Stark *et al.* 2005). At the same time, molecular studies that focused on the *miR-17/92* cluster and its transcriptional regulators revealed intricate feed-back loops formed by TFs and miRNAs (He *et al.* 2005; O'Donnell *et al.* 2005; Sylvestre *et al.* 2007; Woods *et al.* 2007). These indications and others have led to the notion of a global coupling between miRNAs and transcription regulators. This notion was suggested initially by Hornstein and Shomron (Hornstein and Shomron 2006) as a means to provide "canalization" and robustness to developmental programs. The coupling between TFs and miRNAs continued to be a major motivation of research in the coming years (See Discussion and reviewed in (Martinez and Walhout 2009)).

Regulation of cell proliferation

Mammalian cell proliferation is controlled by numerous regulators, among the most important of which are the *E2F* on the one hand, and *p53* on the other. *E2F* is a family of eight TFs, three activators and five repressors of gene expression, which according to our current understanding, bind to the same consensus site. *E2F* activates cell cycle genes, for example *CyclinE* and the *MCM* family, which are crucial for progressing into S phase (G1/S transition) and for DNA replication. However, it was also implicated in the control of the G2/M transition (Ishida *et al.* 2001), mitosis and DNA repair (Polager *et al.* 2002), and it also controls the transcription of many signal transduction genes (Chaussepied and Ginsberg 2005). *E2F* transcriptional activity is coordinated with the different phases of the cell cycle. Consequently, *E2F* target genes have a cell cycle periodic expression pattern. A major part of the tight control of *E2F* levels in normal cells is done by *E2F* itself, which transcriptionally activates a number of its own negative regulators, the most prominent being *pRb* and its family members, the pocket proteins *p107* and *p130*. *pRb*, a well-known tumor suppressor that was discovered

as the mutated gene in Retinoblastoma tumors, represses *E2F* in its dephosphorylated state, by directly binding to it and inhibiting its transcriptional activity. In quiescent cells, the activator *E2F*s are bound to and repressed by *pRB*, while upon growth stimuli, they dissociate and the activator *E2F*s are free to promote cell cycle progression (Polager and Ginsberg 2008).

E2F is well-known for its regulation of cell cycle progression and promotion of proliferation. However in recent years it became evident that it also positively regulates apoptosis. The mechanisms behind the "decision making" of *E2F* to promote either of these two opposite fates are still largely unknown (Ginsberg 2002; Polager and Ginsberg 2008).

The most prominent guardian of proper cell growth and proliferation is the tumor suppressor *p53*. *p53* is famous for its role in protection against cancer, as it is activated by an enormously wide variety of stresses, inducing protective alterations in cell proliferation: either cell cycle arrest, when the sensed damage can be repaired, or apoptosis (Oren 2003) and, as revealed in the recent years, senescence (Brown *et al.* 1997). *p53* exerts its effect on undesired cell proliferation mainly by its transcription factor activity, regulating the expression of hundreds of target genes, which vary under different stresses and different cell fates. The most prominent of *p53*'s targets are *p21* (el-Deiry *et al.* 1993), a major regulator of G1 arrest, and *Mdm2*, an E3 ubiquitin ligase that negatively regulates *p53* protein by marking it to degradation (Haupt *et al.* 1997). *p53* can serve as both transcriptional activator and as a repressor of gene expression. Its transcriptional activation is done through the binding of a *p53* tetramer to a fairly well defined responsive element, whereas its repressive effect on gene expression is achieved via various other mechanisms, both direct and indirect (Vousden and Prives 2009). *p53* is activated in response to a variety of stresses, including DNA damage, oxidative stress, ribosomal stress, spindle damage (Aylon *et al.* 2006) and others (reviewed in (Murray-Zmijewski *et al.* 2008)). It has become evident in the past few years that *p53* serves as a major regulator not only in response to stress, but also throughout the life cycle of normal cells (Vousden and Prives 2009). One intriguing example is that of cellular senescence.

Normal cells enter senescence either after a finite number of cell divisions, termed "replicative senescence", which is said to be the cellular manifestation of aging, or they may senesce in response to various stresses, such as oncogenic stress (reviewed in (Mooi and Peeper 2006)). In fact, over-expression or over-activation of many oncogenes, including *RAS*, *RAF*, *MYC* and others, eventually leads to oncogene-induced senescence (reviewed in (Di Micco *et al.* 2007)). *p53* was shown to be necessary for both replicative and oncogene-induced senescence (Serrano *et al.* 1997).

The *p53* and *E2F* pathways have many points of cross-talk. In fact, both the *pRb-E2F* pathway and the *p53* pathway have been shown to be major regulatory pathways of cellular senescence (Mooi and Peeper 2006). Additionally, even though the regulation of *E2F* promotion of proliferation vs. apoptosis is yet undetermined, one of the proposed mechanisms downstream to its apoptotic activity is via indirect activation of the *p53* pathway. This is done through the transcriptional induction of *ARF* by *E2F*1. *ARF* promotes *p53* stabilization by sequestration of *p53* E3 ligase, *MDM2* (Polager and Ginsberg 2008).

Another regulator of cell proliferation is the *cMYC* oncogene. It was primarily called an oncogene as it was found to be overexpressed in a variety of cancers (Adhikary and Eilers 2005). In normal cells, *cMYC* is rapidly degraded and thus its levels are restrained. *cMYC* is essential for the proliferation, as its knock-out causes embryonic lethality in mice (de Alboran *et al.* 2001), whereas complete knock-out in Mouse Embryonic Fibroblasts (MEFs)causes proliferation arrest (Trumpp *et al.* 2001).

cMYC, as well as its homologs n*MYC*, s*MYC* and l*MYC*, are transcription regulators, and they heterodimerize with *MAX* and bind their target consensus sequence. When *MAX* heterodimerizes with its other partners, *Mad* or *Mnt*, they bind the same consensus elements on DNA and repress transcription, while *MYC* alone cannot bind DNA at all (Adhikary and Eilers 2005). *cMYC* can function both as a transcriptional activator and a repressor. Its most well-known upregulated targets include *CyclinD1* and *2*, *CDK4* and others, which are presumed to contribute to its proliferation promoting activities (Adhikary and Eilers 2005). Interestingly, *cMYC* was shown to have thousands of target

genes, estimated to comprise up to 15% of the genome. Even more intriguing is the fact that mapping of *cMYC* binding sites on DNA showed that *cMYC* binds to many thousands of sites, most of which are intergenic and are very far from known genes. This has led to the hypothesis that *cMYC* might be a global regulator of chromatin opening, and this activity is required for cell proliferation, and when *cMYC* is overexpressed it leads to genomic instability and eventually transformation (Knoepfler 2007).

cMYC is actually transcriptionally repressed by *p53*, although the exact mechanism, whether direct or indirect, is still under debate (Ragimov *et al.* 1993; Ho *et al.* 2005). In turn, *cMYC* is known to transcriptionally repress p21, which allows *cMYC*-transformed cells to overcome G1 arrest following DNA damage (Seoane *et al.* 2002). The *E2F* family member *E2F-2* was found to be a target for transcriptional activation by *cMYC* (Sears *et al.* 1997). *cMYC* is also one of the oncogenes that causes the activation of the oncogene-induced senescence defense mechanism in normal fibroblasts (Drayton *et al.* 2003).

In summary, these three major pathways, dominated by these regulators of cell proliferation, *E2F*, *p53* and *MYC*, are not independent. Rather, they cross-talk at many levels. Whereas *cMYC* is a bona-fide oncogene, and *p53* a prominent tumor suppressor, *E2F* remains somewhat elusive in these definitions, being able to promote both proliferation and apoptosis.

miRNAs in control of cell proliferation and cancer

miRNAs were shown to be differentially expressed in cancers (Lu *et al.* 2005; Rosenfeld *et al.* 2008). Different miRNAs were defined as oncogenes or tumor suppressors, as their overexpression or inhibition were respectively found to occur in human cancers (Zhang *et al.* 2007). A prominent example is the *miR-15a/16-1* cluster, residing in the *DLEU2* (Deleted in LEUkemia 2) non-coding RNA. This genomic region was long known to be frequently deleted in leukemia (Migliazza *et al.* 2000; Migliazza *et al.* 2001), and was shown years later to encode the primary transcript of these miRNAs

(Calin *et al.* 2002). In general, it was demonstrated that the global levels of miRNAs in tumors are lower compared to their normal paired tissues (Lu *et al.* 2005). Nonetheless, specific miRNAs were shown to be overexpressed in certain types of cancers, and were considered to facilitate them, one example being the *miR-17/92* cluster, whose genomic region is amplified in B-cell lymphomas (He *et al.* 2005). miRNAs that are differentially expressed in tumors were shown to target key regulators of proliferation and cancer (reviewed in (Kent and Mendell 2006)).

miRNAs regulation is also relevant directly to cell proliferation (discussed in (Carleton *et al.* 2007; Bueno *et al.* 2008; Chivukula and Mendell 2008)). miRNAs were demonstrated to be transcriptionally regulated by the three major regulators of cell proliferation regulators discussed above: *E2F*, *p53* and *cMYC*, through a complex network. *cMYC* was shown to activate the transcription of the same *miR-17/92* cluster (He *et al.* 2005; O'Donnell *et al.* 2005). Later it was shown that *cMYC* actually induces the widespread repression of a large set of miRNAs in a lymphoma mouse, which contributed to its tumorigenesis (Chang *et al.* 2008). On the other hand, a highly characterized miRNA target of *p53* is *miR-34a*. *miR-34a* was shown to be transcriptionally activated by *p53* (Chang *et al.* 2007; Raver-Shapira *et al.* 2007; Tarasov *et al.* 2007) and to contribute to both apoptosis (Chang *et al.* 2007; Raver-Shapira *et al.* 2007), G1 arrest (Tarasov *et al.* 2007) and senescence (Tazawa *et al.* 2007; Kumamoto *et al.* 2008). Moreover, in line with the cross-talk described above between the *p53* and *cMYC* pathways, and on top of *p53* transcriptional repression of *cMYC*, we recently reported (see results) that *miR-34a* targets *cMYC* for translational inhibition (Christoffersen *et al.* 2009). Another *p53* induced miRNA, *miR-145*, was recently found to repress *cMYC* (Sachdeva *et al.* 2009), again manifesting that *p53* repression of *cMYC* is multifaceted.

E2F was also described as a *miR-34a* target for repression, in the context of colon cancer (Tazawa *et al.* 2007). It was thoroughly described to be involved in a negative feedback loop with the *miR-17/92* cluster (O'Donnell *et al.* 2005; Sylvestre *et al.* 2007; Woods *et al.* 2007).

We further elaborated the view of *E2F*-regulated miRNAs, identifying the *miR-106b/93/25* cluster and its paralogs *miR-106a/363* cluster as additional *E2F* transcriptional activation targets, and showed their wiring in a network of feedback and feed-forward loop with *E2F*, and their relationship to *p53* regulation of cellular senescence in normal fibroblasts (Brosh *et al.* 2008). We also determined that *E2F* is repressed by *miR-106b/93/25*. As stated above, *E2F* regulation of proliferation vs. apoptosis is still a question in debate. One of the possibilities discussed is the threshold theory, assuming that at medium levels *E2F-1* promotes proliferation, whereas when its levels are high it ends up promoting apoptosis. Repression of *E2F*s by miRNAs was assumed to be one of the mechanisms that keeps *E2F-1* levels "below the radar", and thus prevents it from reaching its apoptotic threshold (Chivukula and Mendell 2008).

While it is evident that miRNAs play an important role in the regulation of cell proliferation and cancer, their exact positioning in the network along with other well-known regulators of cell proliferation, is still a subject of extensive investigation. During my PhD I examined the network wiring of miRNAs together with transcription regulators, in an attempt to characterize the network and gain more insight on the regulatory relationship of miRNAs with other regulatory layers in mammalian cells. This has led me to reveal two major principles of miRNA regulatory networks: miRNA combinatorial regulation and the coupling between post-transcriptional regulation by miRNAs and transcription regulation. These principles were further supported by other studies as well (Tsang *et al.* 2007; Zhou *et al.* 2007; Sinha *et al.* 2008; Re *et al.* 2009) (and reviewed in (Martinez and Walhout 2009)). Following the global analysis I further continued to characterize the role of coupling of miRNAs with transcription regulators in the context of cell proliferation and senescence, and establish their importance in the complex network that maintains proper cell growth (Brosh[*], Shalgi[*] et al 2008, and Christofferson et al. 2009). Finally, I developed an interest in the evolution of miRNAs, elaborating on known origins of miRNAs, transposable elements, while also reporting on a new potential origin for novel miRNAs across animal evolution (Dahary[*], Shalgi[*] et al.

2009). Intriguingly, the seemingly divergent interest of evolutionary origins of miRNAs has led to several interesting hypotheses on a global role for miRNA in the maintenance of cell fate and genomic integrity (Shalgi 2009; Shalgi *et al.* 2009).

Published Results

Global and local architecture of the mammalian microRNA-transcription factor regulatory network (Shalgi et al. PLoS Comput. Biol. 2007)

Wishing to explore the possibility of coordination between transcription and post-transcription regulation, we asked whether microRNAs are involved in a cross-talk with the transcriptional program in a mammalian cell, and what implications such multi-level control might have on cell fate.

We explored the network of miRNA regulation in mammalian genomes, in an attempt to provide insights on the network architecture, building blocks, and other global and local properties. Network characterization was done in many other biological networks, mainly in transcription networks, discovering combinatorial regulation (Pilpel *et al.* 2001; Segal *et al.* 2003), identifying network motifs, recurring regulatory architectures that are statistically overrepresented in the network, and provided insights on their function and evolution (Milo *et al.* 2002; Shen-Orr *et al.* 2002; Mangan and Alon 2003). We were among the first to globally investigate the network aspect of miRNA regulation. Using published datasets, (TargetScan and PicTar) of evolutionarily conserved miRNA target site predictions in four mammalian species (human, mouse, rat and dog) (Lewis *et al.* 2003; Krek *et al.* 2005; Lewis *et al.* 2005), we asked how the different layers of transcription regulation and post-transcriptional silencing are integrated in a joint regulatory network. For this purpose we incorporated additional databases including evolutionarily conserved Transcription Factor (TF) binding sites in promoters of protein coding genes and of miRNAs (Karolchik *et al.* 2003).

Global network properties – target hubs and miRNA combinatorial regulation

Looking at the distribution of predicted miRNA – target interactions in the two datasets, we noticed that they are highly non-random, as they show several "miRNA hubs", i.e. miRNAs with relatively high numbers of targets, as well as many genes that

are each targeted by a large number of miRNAs. We named these genes "target hubs", after a recent definition for genes that are regulated by multiple TFs in the yeast transcription networks (Borneman *et al.* 2006). These target hubs, which are at the high tail of the distribution of number of predicted miRNAs per gene, were highly enriched for transcription regulators and for developmentally related genes. In addition, looking at the distribution of their 3' UTR length and the density of miRNA binding sites in their 3' UTR (Stark *et al.* 2005), they were significantly more dense in terms of number of different miRNAs targeting them, compared to the entire gene set examined.

Examining combinatorial regulation within miRNAs, we wanted to identify pairs of different miRNAs that tend to co-regulate large sets of target genes. For this purpose we developed a new statistical method, which takes into account the distribution of in- and out-degrees in the input networks of miRNA-target pairs, and is based on the edge-swapping algorithm by Shen-Orr et al. and Milo et al. (Milo *et al.* 2002; Shen-Orr *et al.* 2002). We identified a network of miRNA co-regulation composed of hundreds of miRNA pairs. This higher-level regulatory network turned out to be scale-free (Jeong *et al.* 2000), as a small number of miRNAs had many co-regulating miRNA partners, while many others were "end nodes", and had only one or two co-regulating partners.

Local network properties – coupling of miRNAs and transcription regulation as a recurrent principle in the network

In order to construct an integrated network of transcription and post transcriptional regulation we looked for pairs of a specific miRNA and a specific TF, which co-regulate a set of common targets: we looked for specific pairs of miRNAs and TFs that co-target a large set of genes, having conserved binding sites on their 3' UTRs and promoters (respectively) of the same genes. We found hundreds of such pairs that are statistically significant. When looking for more complex architectures, asking which of the significant pairs of miRNA-TF partners were also regulating each other, either directly or by a mediator, we found dozens of Feed Forward Loop (FFL) architectures. The different FFLs we identified were found to be over-represented in the network,

suggesting that coupling of transcription and post-transcription regulation by miRNAs is a selected network motif. Strikingly, the regulatory components of these network motifs (namely the miRNAs and the TFs) tend to be either highly correlated or highly anti-correlated in their expression levels across tissues. This agrees well with a suggested role for these network motifs in maintaining robustness of developmental programs (Stark *et al.* 2005; Hornstein and Shomron 2006). The observed tendency of correlations may hint at mechanisms of delayed shut-down in temporal expression in the same tissue, in cases where high positive correlation in tissue expression between the two regulators is observed, and on the other hand, a robust shut-down of genes, facilitating "spatial miRNA-target avoidance" (Stark *et al.* 2005; Baek *et al.* 2008) where TF and miRNA anti-correlate in their expression pattern across tissues.

Together these findings provide new insights into the architecture of the combined transcriptional–post transcriptional regulatory network, and reveal two major principles of miRNA network wiring: the first is combinatorial regulation by multiple miRNAs, and the second is tight coupling with transcription regulators.

The study not only determined general principles, but also provided prioritized lists of predictions of specific miRNA cooperating pairs and specific miRNA-TF pairs that are suspected to co-regulate mutual targets, or form FFL regulatory circuits in the network. Several interesting examples were highlighted, one of which was the predicted FFL of *miR-93* and the *E2F* family of TFs, for which I suggested a possible role in regulation of cell proliferation and cancer. The regulators of this FFL, consisting of *E2F*, a major regulator of cell cycle and proliferation, and *miR-93*, a member of a large miRNA family, whose other members from the *miR-17/92* clusters were already implicated in having an oncogenic role in cancer (He *et al.* 2005), and being a targets of with *E2F* (O'Donnell *et al.* 2005).

Having established globally the coupling of transcription regulation with post-transcriptional miRNA regulation as a recurrent principle in the regulatory network, I wished to substantiate this principle experimentally, and further explore the role of such

particular FFLs in control of cell proliferation. My next study involved high resolution molecular characterization of the *miR-93-E2F* FFL.

p53-repressed miRNAs are involved with E2F in a feed-forward loop promoting proliferation (Brosh, Shalgi* et al. Mol. Syst. Biol. 2008)*

This is a collaborative study, performed together with Ran Brosh from the Rotter lab.

E2F-miR-106b/93/25 FFL

Following my previous work, showing that miRNAs and TFs that form FFLs with common targets are an overrepresented architecture in the mammalian regulatory network, I wished to further study such FFLs experimentally, and to establish their biological role in human cells. One interesting prediction that resulted from my previous work suggested that *E2F* and *miR-93* may cooperate in regulating shared target genes, and also regulate each other, forming a composite FFL (Shalgi *et al.* 2007). As *E2F* is a major cell cycle regulator I suspected that such FFL would be involved in the regulation of cell proliferation. *miR-93* is part of a polycistronic cluster of three miRNAs, *miR-106b*, *miR-93* and *miR-25*, located in close proximity to each other within an intron of the *MCM7* gene, which is a verified target of the *E2F* family (Leone *et al.* 1998). These three miRNAs are part of a family of 15 miRNAs, transcribed from three paralogous polycistronic loci in the genome: the *miR-106b/93/25*, *miR-17/92* and *miR-106a/363* polycistrons (Tanzer and Stadler 2004), and sharing together three miRNA seed sequences. Previous studies have found that the *miR-17/92* polycistron is a target of *E2F* family members (O'Donnell *et al.* 2005; Sylvestre *et al.* 2007; Woods *et al.* 2007), and have a potential role in cancer (He *et al.* 2005). It was therefore plausible that these miRNAs might share similar regulatory architecture with the *miR-106b/93/25* polycistron. Thus I hypothesized that my previous prediction of the *E2F – miR-93* FFL probably involves not only *miR-93*, but also its polycistron members *miR-106b* and *miR-25*, and the rest of the miRNAs in the family as well, and is likely to have a role in regulation of cell proliferation in human cells.

miR-106b/93/25 and paralogs are downregulated by *p53* during senescence in normal cells, and upregulated in breast cancers harboring mutant *p53*

Interestingly, we observed the co-regulation of the superfamily of miRNAs in several experimental systems that relate to normal cell proliferation and senescence. My collaborator from the Rotter lab, Ran Brosh, performed miRNA microarray profiling on WI-38 cells (normal human embryonic fibroblasts) where *p53* was inactivated and cells were grown until they entered replicative senescence (Brosh *et al.* 2008). Clustering analysis of this data revealed an interesting cluster that included many of the family members (in fact, all the ones that were technically detectable in the arrays), whose expression pattern showed a *p53* dependent repression, specifically when cells became senescent. This finding was corroborated by another, independent expression dataset from the Harris lab, where again the family miRNAs were shown to be repressed in senescent WI-38 cells, and also in senescent MRC5 cells, demonstrating that this pattern might be a more general phenomenon, beyond the specific system we used. The same pattern was intriguingly found to be related to breast cancer as well. Clustering analysis of miRNA microarray expression data from a set of human breast tumors of various grades and subtypes (original samples were presented in (Sorlie *et al.* 2006)) revealed that all 15 family members were clustered together, along with other *p53*-repressed miRNA members from the original cluster, and showed an increase in their expression in tumors harboring mutations in *p53*. These findings pointed to the relevance of this family of miRNAs both *in-vitro* for the process of cellular senescence in normal cells and *in-vivo* in breast cancers. This encouraged us to validate the regulatory connections in the proposed *E2F-miR-106b/93/25* and paralogs FFL, further interrogate the mechanism of *p53* repression, and substantiate its role in the context of proliferation and senescence in normal cells.

Validation of the FFL

As all three polycistrons contain conserved *E2F* binding sites in their putative promoter regions, we employed chromatin immunoprecipitation (ChIP) to validate that *E2F-1* indeed binds its sites upstream to all polycistrons; this was done with the help of our collaborators in the lab of Prof. Doron Ginsberg from Bar-Ilan University.

Using a derivative of WI-38 cells stably expressing an inducible *E2F-1* (ER-*E2F-1*), we also showed that *E2F-1* can transactivate the expression of representative miRNAs from all three polycistrons, and that this transactivation is relatively fast (within 4 hours following *E2F-1* induction). We also demonstrated that this is a direct transcriptional activation, as it was not affected by treatment with cycloheximide. Under more endogenous settings, activation of endogenous *E2F-1* through the infection with a recombinant retrovirus encoding the adenoviral *E1a* protein also caused upregulation of these miRNAs.

Next, we cloned the genomic region containing the *miR-106b/93/25* polycistron (the entire intron of the *MCM7* gene) into a miR-Vec expression vector (Voorhoeve *et al.* 2006), and overexpressed the polycistron in WI-38 cells. We saw a reduction in *E2F-1* protein levels, and a mild reduction was also observed in its mRNA level. A similar result was obtained in MCF-10A cells (non-transformed breast epithelial cells).

***miR-106b/93/25* repression by *p53* is mediated via *E2F*1**

Treating WI-38 cells with Nutlin, a small molecule that stabilizes *p53*, we observed a reduction in *E2F-1* protein levels, but to our surprise *E2F-1* mRNA levels were also markedly decreased. This effect was abolished in WI-38 cells where *p53* had been knocked-down by siRNA. Expression of the *miR-106b/93/25* polycistron was downregulated accordingly. To show that the down-regulation was not only correlated, but also dependent on *E2F-1*, we applied Nutlin to WI-38 cells that had been previously infected with the Adenovirus *E1a* protein, which induces endogenous *E2F* activity. We found that the *p53*-mediated repression of the *miR-106b/93/25* cluster was abolished in the E1a overexpressing cells. In addition, when we knocked down *E2F*-1 in WI-38 cells using siRNA, *miR-106b/93/25* levels were decreased, but enhanced stabilization of *p53* (by Nutlin treatment) did not cause a further reduction in the levels of these miRNAs. We therefore concluded that indeed the repression of this miRNA family by *p53* was mediated via *E2F*.

The *E2F-miR-106b/93/25* FFL targets anti-proliferative genes and affects pivotal characteristics of proliferation

Looking at mutual targets, I compiled a comprehensive list of known *E2F-1* target genes from the literature (Ishida *et al.* 2001; Muller *et al.* 2001; Weinmann *et al.* 2001; Ma *et al.* 2002; Polager *et al.* 2002; Ren *et al.* 2002; Stanelle *et al.* 2002), and looked among them for targets of *miR-106b/93/25*, as predicted by PicTar (Krek *et al.* 2005). The set of mutual targets was enriched with cell cycle genes, specifically those that are negative regulators of the cell cycle, and of *E2F* itself. We observed that the protein levels of several candidates, such as the pocket proteins *pRb* and *p130* and the *CDK* inhibitors *p21* and *p57*, were down-regulated in WI-38 cells over-expressing *miR-106b/93/25*, while their mRNA levels remained unchanged. To further substantiate the nature of the mutual targets at a more global level, I looked at the expression data of Milyavsky et al. (Milyavsky *et al.* 2005), where primary WI-38 cells were gradually manipulated in vitro and gained an accelerated proliferation phenotype until they became fully transformed into tumorigenic cells. Specifically, I considered the expression profile of genes that were heavily targeted by the family miRNAs as well as by other *p53*-repressed miRNAs. Surprisingly, I saw that most of the targets were anti-proliferative, as exemplified by their expression profile which was decreased when cells gained the accelerated proliferation phenotype. In addition, promoter analysis revealed that their promoters were enriched with *E2F* binding sites, manifesting the coupled transcriptional-post transcriptional regulation of these genes by *E2F* and the *miR-106b/93/25* family. Finally, we tested the effect of perturbation of the FFL on the fate of WI-38 cells. Over-expression of the *miR-106b/93/25* polycistron caused an enhanced proliferation phenotype, as exemplified by increased proliferation rates, a significantly larger fraction of cells in S phase, and enhanced colony formation rates. In addition, similarly to *p53* inactivation, such overexpression caused a substantial delay in cellular senescence.

In summary, we revealed the role of the FFL of *E2F* and *miR-106b/93/25* polycistron and paralogs in regulation of cell proliferation. Among the targets of this polycistron are some key players in negative regulation of cell proliferation. These

miRNAs were indeed shown to be up-regulated in breast cancer tumors, indicating that they might indeed be associated with cancer *in-vivo*. We established that in normal cells, the repression of the FFL regulators by *p53* is necessary for the appropriate senescent fate, and perturbations in the FFL result in phenotypic changes that alter cell fate and exhibit tumorigenic nature.

p53-independent upregulation of miR-34a during oncogene-induced senescence represses cMYC (Christoffersen et al. Cell Death Differ 2009)

This study was done in collaboration with the Anders Lund lab in the University of Copenhagen, Denmark.

In this collaborative study, we investigated the role of miRNA regulation and its coupling with transcription regulation, in oncogene-induced senescence. For this purpose, our collaborators used the model of *B-RAF* overexpression in normal fibroblasts, which is known to induce senescence (Zhu *et al.* 1998). Specifically, following 3-4 days of constitutive activation of the *B-RAF* oncogene in normal fibroblasts, (TIG3 cells), they enter senescence. miRNA expression was assayed following this oncogene-induced senescence, and *miR-34a* was found to be the most prominently induced miRNA. Wishing to discover the potential targets of *miR-34a* during *B-RAF* induced senescence, *miR-34a* was inhibited using a *miR-34a*-specific locked nucleic acid (LNA) oligonucleotide in control TIG3 cells and in cells following *B-RAF* activation. Global mRNA expression was then assayed, with the intention of identifying changes in the mRNA levels of relevant *miR-34a* targets. This method had been used before to identify targets for *miR-122* and *miR-21* (Krutzfeldt *et al.* 2005; Frankel *et al.* 2008).

Global analysis of mRNA expression following *miR-34a* inhibition during *B-RAF*-induced senescence predicts *cMYC* as a potential mediator of *miR-34a* regulation

Preliminary analysis of global mRNA expression showed that *B-RAF* activation resulted in major changes in gene expression, while *miR-34a* inhibition had a relatively minor effect. For validation purposes, we verified that *miR-34a* inhibition resulted in de-repression of several previously reported *miR-34a* targets, including *BCL2* and *CDK6*, and, to a lesser extent, *MET* and *CCND1* (Bommer *et al.* 2007; He *et al.* 2007; Sun *et al.* 2008). However, on a global scale, the overall ensemble of predicted mRNA targets of *miR-34a* did not display a *miR-34a* dependent effect on expression, and no enrichment of *miR-34a* seed sequences was found among the differentially expressed mRNAs.

To highlight changes dependant on *miR-34a*, I focused on the top 20% of the genes that were most influenced by *miR-34a* inhibition, and subjected them to clustering analysis: the top 5% most upregulated and 5% most downregulated genes in each pair of samples were selected (the *B-RAF*/*miR-34a* LNA vs. *B-RAF*/Scramble samples, and the Cont/ *miR-34a* LNA vs. Cont/ Scramble samples), and the four lists were unified, resulting in 1730 genes.

Although here too, no enrichment of motifs matching the *miR-34a* seed sequence was detected within the 3' UTRs of genes in clusters resulting from this dataset, several clusters displayed an interesting *miR-34a*-dependency. Focusing on a cluster of ~350 transcripts, which were repressed upon *B-RAF* activation, but the inhibition of *miR-34a* alleviated this *B-RAF*-mediated repression, I turned to further characterize it.

Functional annotation analysis using DAVID (Dennis *et al.* 2003) revealed a significant enrichment for genes related to the cell cycle ($p<1.9\times10^{-11}$). Of note, CDK6 and BCL2, two of the known targets of *miR-34a*, were included in this cluster. Since most of the transcripts in this cluster did not contain target binding sites for *miR-34a* in their 3' UTRs, I suspected that a common transcriptional regulator, which is a *miR-34a* target, might be responsible for the observed expression pattern, thereby mediating a global *miR-34a* effect. Interestingly, when subjecting the promoters of the genes in this cluster to a motif finding search (using AMADEUS (Linhart *et al.* 2008)), an enrichment for a motif resembling a *cMYC* binding site ($p<3.3\times10^{-12}$) was found, suggesting that *miR-34a* might influence gene expression via targeting *cMYC*.

Inspection of the 3'UTR sequence of *cMYC* mRNA revealed the presence of a perfectly complementary and evolutionarily conserved 7-nt match to the seed region of the two other members of the *miR-34* family, namely *miR-34b* and *miR-34c* (Lewis *et al.* 2003) and a 6-nt seed match to *miR-34a*. *miR-34c* was recently suggested to target *cMYC* mRNA directly, on the basis of a reporter assay (Kong *et al.* 2008), but in the TIG3/*B-RAF* experimental system neither *miR-34b* nor *miR-34c* were detectable above background, suggesting that they do not contribute to changes in gene expression in these cells. Of note, whereas *cMYC* protein levels were found to be reduced during oncogene-induced senescence, no change was observed in *cMYC* mRNA levels following *B-RAF*

induction or *miR-34a* overexpression (as evident from the microarray), suggesting post-transcriptional regulation, possibly at the level of translation.

To validate the prediction, I transfected H1299 (lung cancer) cells with *miR-34a* LNA inhibitors and observed an upregulation of *cMYC* protein levels. Here, too, *cMYC* mRNA levels were not affected by *miR-34a* inhibition. In parallel, our collaborators verified the same behavior in TIG3 cells, and showed that during oncogene-induced senescence, *miR-34a* contributes to the repression of *cMYC* protein.

Taken together with the fact that no enrichment of *miR-34a* seed/motif was found among *miR-34a* affected genes, this may indicate that *miR-34a* affects its direct targets primarily through inhibition of translation and not through mRNA degradation, and the results of the microarray demonstrate mainly indirect regulation by *miR-34a*.

In this study we have shown another aspect of coupling between transcription and post-transcriptional control, and its relevance to oncogene-induced senescence. Whereas *p53* was shown to transcriptionally activate *miR-34a* (Chang *et al.* 2007; He *et al.* 2007; Raver-Shapira *et al.* 2007; Tarasov *et al.* 2007), also in the context of senescence (Kumamoto *et al.* 2008), in our collaborative study the TF *ELK-1* was shown to be a major player in *miR-34a* upregulation. In addition, *p53* has been known for years as a repressor of *cMYC* transcription (Ragimov *et al.* 1993). A recent study reported that another miRNA, *miR-145*, is transcriptionally induced by *p53* and represses *cMYC* (Sachdeva *et al.* 2009). In our study we contributed to understanding the role of *p53* and other transcriptional networks controlling gene expression during oncogene-induced senescence, which spreads its regulation through post-transcriptional repression by miRNAs. We demonstrated how, through regulation of *cMYC* by *miR-34a*, the regulatory effect is propagated to hundreds of target genes.

A likely evolutionary origin of microRNAs in animals (Dahary, Shalgi* et al. 2009, Submitted)*

This study was done in collaboration with Dvir Dahary from the Pilpel lab.

The dramatic increase in morphological complexity in animal evolution has recently been attributed to non-coding RNAs (ncRNAs) and specifically to miRNAs, postulating that they represent a principal layer of gene regulatory networks in metazoans (Sempere *et al.* 2006; Niwa and Slack 2007; Grimson *et al.* 2008). Recent studies revealed several episodes of expansion in the number of miRNA families, which correspond to the introduction of novel morphological features during animal evolution (Sempere *et al.* 2006; Niwa and Slack 2007; Grimson *et al.* 2008). The three major "bursts" of miRNA innovations correlate with the evolutionary branching of vertebrates, mammals and primates. However, despite the enormous potential role ascribed to miRNAs in evolving animals complexity, their origin, i.e. the evolutionary mechanism that gave rise to them, is not yet fully understood.

In this collaborative study we explored two major origins for miRNA innovations during evolution, and examined the different dynamics of miRNA innovations along the evolutionary tree with respect to these two origins. The first origin is transposable elements, suggested before as an origin for miRNAs by several studies (Piriyapongsa and Jordan 2007; Piriyapongsa *et al.* 2007). The second is a newly suggested origin: CpG rich regions, and in particular – CpG islands.

We hypothesized that miRNAs were evolutionarily born from genomic clusters of CG dinucleotides, termed 'CpG islands' (CGIs). We suggested that these unique regions could serve as a template for novel hairpins, and hence miRNAs, for several reasons. First, the basic structural feature of animal miRNAs, a long stable hairpin, requires a relatively high GC-content, which is especially crucial as most of the genome is generally

AT-rich due to the general mutational transition bias towards AT (Marais 2003). Moreover, CGIs possess basal promoter activity (Sandelin *et al.* 2007). Therefore, such regions might provide the basic requirements for a new miRNA to evolve: being transcribed, and giving rise to a stable hairpin, which may then be processed into a mature miRNA. Such activity may then lead to the selection of the miRNAs during evolution

We examined the localization of all the known miRNAs in the human genome, and observed that as much as 10% of all human miRNAs physically reside within an annotated CGI. If we examine miRNA families, assuming that one ancestral miRNA gave rise to the entire family by paralogous duplications, 12% of miRNA families are CGI-associated. Controlling for possible bias as miRNAs are often transcribed as polycistrons and lie in close proximity to each other on the genome, we still observed more than 12% of miRNA clusters are CGI-associated. This overlap is highly statistically significant, as the frequency of CGI in the human genome is less than 1%. We showed that this over-representation is highly significant even when comparing it to a variety of different background models which take into account other constraints that may occur on miRNA genomic localization, demonstrating that the observed phenomenon is significant by itself and not as a by-product of a third-party genomic feature.

Examining the overlap of human miRNAs with annotated repeats, either LINE, SINE, DNA repeat or LTR, we observed 149 miRNAs that overlap a repeat. Surprisingly, this is a higher fraction (22%) than was previously reported (12%) when previous versions of the human miRNA collection were examined (Piriyapongsa *et al.* 2007).

We then turned to compare the evolutionary origins of interest, i.e. repeats vs. CGIs. We saw that they are completely separate, and there are no miRNAs that reside next to both CGIs and repeats. Checking the entire genome, we indeed saw that TEs rarely overlap with CGIs. But the complete separation of the two groups indicates that these are two separate sources of continuous supply of novel miRNAs during evolution.

Investigation of the evolutionary dynamics of these two potential origins of miRNAs revealed an intriguing picture. We compared the origins of human miRNAs, after we classified them according to their evolutionary age (for simplification, we classified them into four age groups: Primate specific, mammal specific, vertebrate specific and older). We found that almost 40% of human-specific miRNAs overlap a repeat, whereas only 20% out of the mammalian specific miRNAs overlap such repeats, and as expected, almost zero of the vertebrate miRNAs, and zero of the older miRNAs are associated with repeats Examination of CGI-originated miRNAs revealed the opposite trend: ~9% of the primate specific human miRNAs are CGI-associated; this fraction is slightly increased in mammalian and vertebrate specific miRNAs to ~11%, and strikingly, more than 20% of the old miRNAs are CGI overlapping.

Taking into account that the definition of CGI uses somewhat arbitrary thresholds, we also compared the distribution of CpG observed/expected values in the regions of miRNAs of different lineages. Here too we observed that the distribution of CpGs in old miRNAs is markedly shifted to the higher values compared to the others. Even when we filtered out all the CGI-associated miRNAs from all lineages, we still observed a distribution with higher CpG observed/expected values in the old lineages, indicating traces of more additional miRNAs that might have originated from a CGI that was subsequently decayed.

These two opposite trends may suggest that the two separate genomic entities, repeats on the one hand and CGIs on the other, served as suppliers of novel miRNAs during evolution. We suggest that miRNAs are constantly born from repeats and from CGIs and are selected for and thus retained during evolution. However, the CGI-originated miRNA are more likely to be retained in evolution, whereas repeat-originated miRNAs are rapidly born and also decay at high rates along with the repeats that host them. However, it is possible that repeat-originated miRNAs are retained in evolution while the repeats that surround them are actually lost, and therefore some of the vertebrate-specific and older miRNAs were actually originated from repeats. However some evidence point to the other option. In fact, when we compared the numbers of

repeat-originated miRNAs in the primate and mammalian lineages, we see that 108 primate-specific miRNAs are TE-originates, corresponding to 68 novel miRNA families, whereas only 34 mammalian specific miRs are TE-originated, corresponding to 31 miRNA families: more than three times less, or more than twice less when we consider miRNA families. However, when we counted the total sequence of all mammalian and primate repeats which ever gave rise to miRNAs, we see that while mammalian repeats occupy ~1Gbps of the human genome, the primate repeats we counted occupy only ~0.3GBps. Given that mammalian repeats are capable of serving as origins for primate specific miRNAs, the potential of repeat sequence that could theoretically give rise to primate specific miRNAs is the sum of the two, 1.3Gbps. Thus, we see that while mammalian repeat sequences that were retained in the human genome are only 1.3 less than the total TE-sequence in the human genome. This may indicate that indeed miRNAs are generated from TEs across evolution, but disappear much more rapidly than expected by the selection forces that are applied on TEs, and support our model. Extrapolating from these numbers, we may guess that the current collection of human miRNAs is still under selection, and given these rates, out of the current 108 TE-derived human specific miRNAs, only ~41-45 will eventually be further retained in evolution.

Discussion

In my PhD I explored general principles of regulatory networks of miRNAs, and showed how these principles are manifested in the control of cell proliferation and senescence. Initially I demonstrated two themes in miRNA regulatory networks: the first is combinatorial regulation by miRNAs, and the second if the coupling between transcription regulation and miRNA regulation. I showed that both are common motifs in mammalian regulatory networks, and gave a prioritized list of predictions of such cases of coupling between specific TFs and specific miRNAs. These global findings, which were also supported by subsequent studies in the literature, have led me to test experimentally the motifs I predicted, and pursue an investigation of their biological function and significance, in particular in the context of cell proliferation. I continued to characterize in-depth the FFL that consists of E2F and miR-106b/93/25 cluster and family, in what turned out as a fascinating biological manifestation of the above two principles, in the control of cell proliferation and senescence. Our study revealed an intricate network, whereby the E2F-1 - miR-106b/93/25 FFL serve to co-regulate multiple anti-proliferative target genes while also regulating each other. Additionally, the FFL regulators are repressed by p53 specifically when normal fibroblasts enter senescence, and suggested a role for it in breast cancers harboring p53 mutations. We also showed that perturbation in this FFL, by means of over-expression of the miRNAs, bears a detrimental effect on cell fate of normal fibroblasts, granting them with tumorigenic features. Thereafter I continued to explore the role of other cases of coupling of transcription and miRNA regulation in more collaborative studies, and specifically showed the role of miR-34a in oncogene-induced senescence by targeting *cMYC*. Finally, I pursued another interest I had in evolution of miRNAs. In a latest study we describe two alternative modes for miRNA innovations during evolution, elaborating on the existing knowledge on TE-derived miRNAs, and proposing a new origin – CpG Islands. Integration of my findings on principles of miRNA regulation, their role in control of cell proliferation, together with the question of miRNA evolutionary origins have lead me to gain new insights on each of these topic separately, and also on the significant impact of

miRNAs on normal cell fate, and to propose a new notion about their involvement in maintaining genomic integrity, as outlined below.

miRNAs are key regulators of cell fate

miRNAs have emerged in the past decade as important players in numerous cellular and organismal processes in animals and plants (Bartel 2004). Deletion of the *Dicer* gene, encoding the critical enzyme involved in miRNA processing and maturation, is embryonic lethal in both mice (Bernstein *et al.* 2003) and zebrafish (Wienholds *et al.* 2003). Accordingly, many studies showed, using conditional elimination of *Dicer*, that miRNAs are crucial for the proper spatiotemporal development of various tissues and organs ((Lander *et al.* 2001; Bernstein *et al.* 2003; Kanellopoulou *et al.* 2005; Volinia *et al.* 2006; Murchison *et al.* 2007; Chen *et al.* 2008; Damiani *et al.* 2008) and reviewed in (Bushati and Cohen 2007)). Further, mouse embryonic stem (ES) cells defective in miRNA processing were shown to proliferate slower (Murchison *et al.* 2005), and to be impaired in their ability to differentiate (Kanellopoulou *et al.* 2005). In parallel, other studies have shown a major role for miRNAs in development, indicating that many miRNAs are upregulated during the process of ES cell differentiation ((Houbaviy *et al.* 2003) and reviewed in (Gangaraju and Lin 2009)). Numerous miRNAs also play a role in differentiation processes in the adult organism, including in hematopoiesis (Lu *et al.* 2008) and in the germinal center response (Thai *et al.* 2007). In fact, the first miRNAs to be discovered, *lin-4* and *let-7* in *c.elegans*, regulate epithelial cell differentiation (Ambros and Horvitz 1984; Lee *et al.* 1993). In addition, manipulations of individual miRNA genes were shown to result in marked defects at the organismal level (Hornstein *et al.* 2005; Ventura *et al.* 2008; Bonauer *et al.* 2009) (also reviewed in (Smibert and Lai 2008)). Based on these accumulated observations it is plausible to suggest that in many cases miRNAs are indeed a part of the driving force of differentiation processes. miRNAs were also shown to regulate many cellular processes (Alvarez-Garcia and Miska 2005; Erson and Petty 2008), such as cell growth and proliferation (reviewed in (Bueno *et al.*

2008; Chivukula and Mendell 2008)) and apoptosis (reviewed in (Jovanovic and Hengartner 2006)). It appears, therefore, that miRNAs are crucial players in the regulation and determination of cell fate.

Coupling transcriptional and post-transcriptional miRNA regulation in the control of cell fate

Overall, there seems to be a discrepancy between the subtle effect of miRNA on protein levels that has been reported lately in two systematic studies (Baek *et al.* 2008; Selbach *et al.* 2008), and the fact that their effects on cell fate are so profound. Coupling of transcriptional regulation of genes with post-transcriptional regulation by miRNA may facilitate the significant influence miRNAs have n maintaining proper cell fate

One answer to the above discrepancy might argue that the multiplicity of miRNA targets and the simultaneous down-regulation of many proteins have a cumulative effect, eventually exerting a significant impact on cell fate, even though individual proteins are repressed to a very modest extent. This is a valid argument, particularly since some miRNAs were predicted and shown to have multiple targets within the same pathway (Stark *et al.* 2003; John *et al.* 2004; Brosh *et al.* 2008), thus potentially having greater effects on entire pathways than on individual proteins.

Another possible explanation for their significant influence on cell fate may lie in the level of the regulatory networks that miRNAs take central part in. miRNAs do not act in isolation, but rather they regulate target genes combinatorially with one another, and are often embedded within intricate regulatory networks together with TFs. In fact, in my study (Shalgi *et al.* 2007), I demonstrated a tight coupling between post-transcriptional regulation by miRNAs and the regulation of transcription by TFs at the network level. This principle was also independently shown to be a global feature of regulatory networks by several other labs ((Tsang *et al.* 2007; Sinha *et al.* 2008) also reviewed in (Martinez and Walhout 2009). In particular, I showed that in many cases the same TF controls the transcription of both a miRNA and its targets, or is regulated by the same

miRNA with which it shares common targets, forming a diversity of combined transcriptional/post-transcriptional Feed-Forward Loop (FFL). Collectively, such FFLs potentially regulate thousands of target genes. Network analysis showed that these FFLs constitute over-represented architectures in the mammalian regulatory network (Shalgi *et al.* 2007; Tsang *et al.* 2007; Sinha *et al.* 2008).

Network FFLs, initially described by Alon and colleagues, were shown to comprise a major component of the transcription networks in bacteria and yeast (Milo *et al.* 2002; Shen-Orr *et al.* 2002). The discovery that miRNAs and TFs also constitute FFLs offered new possibilities for potential functions for these regulatory units. Clues for the existence of coupling between transcription and miRNA regulation emerge from a very intriguing concept, called miRNA-target avoidance. Two parallel studies, one in *Drosophila* and the other in mammals, have shown that during development as well as in adult tissues, miRNA targets often avoid being expressed in the same tissue, or at the same developmental time, as their potential inhibitory miRNA (Farh *et al.* 2005; Stark *et al.* 2005). In *Drosophila*, it was demonstrated for several cases that a miRNA and its targets are expressed in adjacent tissues during development, or in consecutive developmental stages, and that miRNAs serve as key players in the precise definition of spatiotemporal differentiation boundaries (Stark *et al.* 2005). This phenomenon was observed also in adult tissues and organs in both *Drosophila* (Stark *et al.* 2005) and mouse (Farh *et al.* 2005). Moreover, both studies indicated that this mutual exclusion of miRNAs and their targets does not stem from target degradation by the miRNA. These two studies are in agreement with the notion that post-transcriptional regulation by miRNAs is somehow coordinated with transcription. However, it was not shown originally how, at the mechanistic level, such "miRNA-target spatio-temporal avoidance" is achieved. Combined transcriptional/post-transcriptional FFLs, where the same TF regulates the transcription of both a miRNA and its target genes, or where the miRNA targets a TF and its target genes as well, could serve just that purpose. Such FFLs are thus suggested as a simple mechanism that may facilitate the miRNA-target avoidance phenomenon, where a TF that activates the target genes also represses the transcription of the miRNA in the tissues in which it is expressed, or the miRNA represses both the TF and its target genes,

thereby indirectly causing reduced transcription of its targets in the tissue where it is expressed (Shalgi *et al.* 2007). In addition, such FFLs were further suggested to enable the "canalization" and the fidelity of developmental processes in general (Hornstein and Shomron 2006).

More recently, evidence has been accumulating that such combined transcriptional post-transcriptional FFLs indeed act as functional units in the regulation of cell fate in many cell types and systems (O'Donnell *et al.* 2005; Woods *et al.* 2007; Brosh *et al.* 2008; Marson *et al.* 2008; Cohen *et al.* 2009; Li *et al.* 2009; Sachdeva *et al.* 2009). One striking example, recently published by Marson *et al.* (Marson *et al.* 2008), demonstrated that miRNAs and TFs are involved together in FFLs controlling the maintenance of mouse embryonic stem (ES) cell identity. Consistent with the studies mentioned above (Bernstein *et al.* 2003; Wienholds *et al.* 2003; Kanellopoulou *et al.* 2005; Murchison *et al.* 2005), which showed that complete miRNA ablation from ES cells eliminates their differentiation capacity, Marson *et al.* showed that several FFLs involving miRNAs and ES cell TFs act to regulate ES cell identity and differentiation. For example, the *miR-290-295* polycistronic cluster, containing the most abundantly expressed miRNAs in mouse ES cells, is positively regulated by the ES cell TF *Oct4*, whereas its promoter is co-occupied by *Oct4*, *Sox2*, and Nanog. In addition, *miR-290-295* co-regulates mutual target genes along with these same TFs. Intriguingly, while miR-*290-295* is a rodent specific cluster, a similar FFL involving *Sox* and *Oct4* was computationally predicted in humans in my original study ((Shalgi *et al.* 2007)). This predicted FFL in humans comprises *miR-302*, which shares the same seed as the rodent-specific *miR-290-295*, and was shown to be highly expressed in human ES cells (Laurent *et al.* 2008), perhaps serving as a *miR-290-295* human ortholog. Consideration of these results in the perspective of previous studies on the role of miRNAs in ES cell differentiation supports the conjecture that miRNA-involving FFLs play an important function in this context, and suggests potential conserved roles for similar FFLs in the maintenance of human ES cell identity as well.

My work on miRNA-TF FFL, in collaboration with Ran Brosh from Prof. Rotter's lab, provides yet another perspective on miRNA-TF FFLs involvement in regulation of cell proliferation (Brosh *et al.* 2008). In our study, a family of 15 homologous miRNAs transcribed as three polycistrons: *miR-106b/93/25*, *miR-17-92* and *miR-106a-363*, were shown to form a proliferation-promoting FFL together with the transcription factor *E2F*, targeting a battery of anti-proliferative target genes. Most importantly, we demonstrated that in normal fibroblasts *p53* inhibits this FFL as a central step towards cellular senescence, and when this inhibition is perturbed, normal cell fate is altered: proliferation is accelerated and senescence is delayed. In agreement with these results, breast cancer tumors bearing mutated *p53* showed an elevation in the levels of these miRNAs and were characterized by a high tumor grade, hinting at the role of these miRNAs in promoting proliferation and aggressiveness also *in vivo* in tumors. The same miRNA family was indeed reported in several independent studies to be related to promotion of cancer (He *et al.* 2005; Brosh *et al.* 2008; Li *et al.* 2009) (also reviewed in (Mendell 2008)). Our study illustrates how deregulation of the entire FFL may contribute to aberrant proliferation. It also provides molecular support to another concept of network wiring of miRNAs, which was initially introduced by my earlier computational study (Shalgi *et al.* 2007), namely combinatorial miRNA regulation. In this case our findings reveal combinatorial regulation by family-related miRNAs. Combinatorial regulation by miRNAs was globally predicted based on co-occurrence of miRNA target sites in common gene sets (Shalgi *et al.* 2007), and was also observed experimentally by others later on (Ivanovska and Cleary 2008). miRNAs can be grouped by mature sequence similarity into miRNA families. In some cases, as in the case of the *miR-106b/93/25* family mentioned above, these families are shown to represent paralogous groups of miRNAs of a common evolutionary origin (Tanzer and Stadler 2004). Just as paralogous genes were duplicated during evolution but retained some degree of sequence similarity, these paralogous miRNAs share similarity in their sequence, which immediately suggests that they might also share common target genes. More intriguingly, it seems that in many cases such families had not only retained similar targets, but also retained similar transcriptional programs. As described by our study (Brosh *et al.* 2008), the above family

of 15 miRNAs indeed retained their joint transcriptional regulation by *E2F* and *p53*. Coordinated transcriptional regulation of a family of miRNAs that share the same targets, all of which are part of the same pathway (in this case negative regulators of proliferation), may have a cumulative effect on the overall levels of proteins in the pathway, thus resulting in a strong overall effect on cell fate despite relatively minor effect at the level of a single miRNA and its target.

Coordinated regulation of a family of miRNAs was also shown in other cases (Zhao *et al.* 2005; He *et al.* 2007). For example the *miR-34* family, consisting of two transcription units and three mature family members, were all shown to be transcriptionally activated by *p53* and to contribute to cell cycle arrest (He *et al.* 2007). Moreover, this family was also demonstrated to have shared targets: we discovered that *miR-34a* targets *cMYC* (Christoffersen *et al.* 2009) while others reported that *miR-34c* targets *cMYC* (Kong *et al.* 2008). Additional support to the notion of common regulation of miRNA families was given in both mouse and human ES cells, where several related miRNA families, often sharing similar seeds, were shown to be co-expressed (Laurent *et al.* 2008; Marson *et al.* 2008). Finally, a recent study actually addressed this question directly, verifying that miRNAs from the same family indeed have many shared targets (Ivanovska and Cleary 2008).

Overall, combinatorial regulation by miRNAs, particularly miRNAs from the same family, and shared transcription programs for such miRNAs and their common targets, portray an intricate network architecture, which was both predicted by me as a principle in regulatory networks and demonstrated experimentally by us and others, in several biological systems. Such architectures are not only over-represented by computational means, but may also cumulatively generate a strong output that is likely to account for the observed effects on cell fate and for the alteration of cell fate when the miRNAs are mis-regulated.

miRNAs –guardians of genomic integrity?

The work by Lu *et al.* (Lu *et al.* 2005) was one of the first studies that carried out global expression profiling of miRNAs across a large set of tumors, demonstrating that miRNA expression profiles can be used to classify human cancers of unknown origin. In addition, that study made an interesting observation: in general, tumors have lower levels of miRNAs than normal tissues. The authors suggested that the observed low global levels of miRNAs may be a reflection of the de-differentiated state of tumors. An alternative hypothesis would be that tumors, during the course of cancer progression, evolve to silence the miRNA pathway. In other words, globally avoiding regulation of gene expression by miRNAs may be one of the many ways of cancer cells to enhance their proliferation and tumorigenic potential.

Several lines of evidence support the idea that proliferating cells and cancer cells in particular, find many different ways to avoid post-transcriptional regulation by miRNAs. Some of these mechanisms are straightforward, and are in agreement with the current view on tumor suppressors and oncogenes. For example, the *cMYC* oncogene was found in a lymphoma mouse model to mediate widespread repression of a large set of miRNAs, contributing to tumorigenesis (Chang *et al.* 2008). Other mechanistic possibilities for tumors to avoid post-transcriptional regulation by miRNAs include epigenetic silencing, mutations and deletions of genomic loci encoding for miRNAs (Calin *et al.* 2002; Calin *et al.* 2004; Bueno *et al.* 2008; Datta *et al.* 2008; Lujambio *et al.* 2008; Zhang *et al.* 2008). Another newly described mechanism is the interruption of the miRNA biogenesis pathway, by processes such as nuclear retention of pre-miRNAs (Lee *et al.* 2008), or pri- and pre-miRNA processing blockage, as in the case of inhibition of maturation of the let-7 family by *Lin28* protein (Heo *et al.* 2008; Newman *et al.* 2008; Viswanathan *et al.* 2008). *Lin28* was further shown to promote cancer, and this was attributed to its repression of the *let-7* family (Viswanathan *et al.* 2009). Moreover, a recent report implicates *p53* in the enhancement of miRNA maturation for several miRNAs following

DNA damage (Suzuki *et al.* 2009), attesting to global miRNA upregulation as a possible anti-cancer mechanism. Additional evidence indicates that proliferating cells tend to employ alternative polyadenylation and alternative splicing in order to express mRNAs with shorter 3' UTRs, which have fewer miRNA binding sites (Sandberg *et al.* 2008). It seems that even though certain miRNAs can act as oncogenes when aberrantly expressed, the emerging picture is that, overall, cancers find many ways to globally avoid miRNA-mediated repression. This leads to the emergent question of what is the advantage that miRNA avoidance provides to cancer cells.

The most striking evidence in support of this apparent 'miRNA avoidance' strategy played by tumors is provided by two seemingly contradictory studies, one focusing on cancer cells and the other on normal cells. The study by Kumar *et al.* (Kumar *et al.* 2007) reported that the ablation of miRNAs in various cancer cell lines, using knock-down of *Dicer*, resulted in enhanced cellular transformation, evident by increased colony formation efficiency *in vitro* and increased tumor burden *in vivo*. On the other hand, Mudhasani *et al.* (Mudhasani *et al.* 2008) showed that the total elimination of miRNAs using conditional *Dicer* knock-out results in premature senescence in normal MEFs. This effect was also apparent at the level of the organism, as the knock-out of *Dicer* in keratinocytes and skin epidermis of adult mice resulted in senescence-induced hair loss and skin aging (Mudhasani *et al.* 2008). In these two studies, the same event could lead to two opposite outcomes, depending on the cellular context. Similarly, the activation of oncogenes, such as *RAS*, is one of the hallmarks of cancer, and when occurring in cancer cells will cause the enhancement of their cancerous phenotype. However, in normal cells, oncogene activation will often lead to genomic instability, which is sensed by the DNA damage checkpoint, and leads to *p53* and *Arf*-dependent oncogene-induced senescence (Serrano *et al.* 1997). Importantly, the phenomenon described by Mudhasani *et al.* (Mudhasani *et al.* 2008) was not a classical case of oncogene-induced senescence, as it was not accompanied by the up-regulation of the oncogenes *MYC* or *RAS*, two well-known activators of oncogene-induced senescence, even though both are documented miRNA targets (Johnson *et al.* 2005; Christoffersen *et al.* 2009; Sachdeva *et al.* 2009). Interestingly, however, senescence following *Dicer* knock-out was also mediated through

activation of the p19^{ARF} and *p53* pathways (Mudhasani *et al.* 2008). In fact, the study demonstrated that the depletion of miRNAs led to DNA damage, as evident by *γH2AX* staining, and consequently, through activation of the *Arf* and *p53*-dependent DNA damage checkpoint, resulted in premature senescence. Therefore, in this case too, the same event of global miRNA depletion induces the DNA damage checkpoint in normal cells due to proper *p19^{ARF}* and *p53* activation, while in cancer cells, where these checkpoint response pathways are frequently inactivated (Hanahan and Weinberg 2000), it leads to genomic instability and enhanced transformation.

Following the examination of these two studies (Kumar *et al.* 2007; Mudhasani *et al.* 2008), I would like to raise the suggestion that miRNAs are essential for maintenance of genomic integrity in normal cells. Further, global miRNA inhibition contributes to cancer formation not only by enhancing proliferation, but also by leading to genomic instability, causing increased DNA damage, and thus potentially resulting in increased mutation rates. In principle, miRNAs can therefore be considered as one of the "guardians" of genomic integrity, serving as another regulatory barrier whose removal may be part of a series of events that ultimately lead to cancer.

Involvement of miRNAs in the maintenance of genomic integrity by global repression of transposable elements - a newly proposed role for miRNAs in the nucleus

The next question that may follow is how do miRNAs serve as guardians of genome integrity and cell fate?

One answer may again lie in the level of the network wiring of miRNAs within the complex regulatory network of the cell, as discussed in the previous sections. The coupling of miRNAs with transcription regulation, its combinatorial nature, and its tendency to influence multiple genes in the same pathway may contribute to the great importance of miRNAs in the maintenance of genomic integrity, as it contributes to the fidelity of cell fate. An additional explanation may be based on the conjecture that in

human cells miRNAs are involved not only in post-transcriptional silencing, but also in repression at the level of the DNA.

As outlined above, the ablation of *Dicer* in several normal fully differentiated tissues and cell types resulted in DNA damage, triggering the DNA damage checkpoint and the subsequent chain of events, which operated normally (Mudhasani *et al.* 2008). Thus, the major defect in these normal cells was the DNA damage itself, while the subsequent downstream pathways (the sensing of DNA damage, activation of *p53* and *Arf*, and subsequent induction of premature senescence) were intact. I would like to suggest the hypothesis that this DNA damage has not occurred through changes in protein expression due to the absence of one or several specific miRNAs. Indeed, cancers find many ways to globally repress and avoid silencing by miRNAs, and this is not restricted to one or several specific miRNAs (as discussed above). It is plausible that if miRNA ablation results in genomic instability, miRNA avoidance mechanisms will be selected for during cancer progression in order to increase the cells mutation rates, which will further contribute to tumor evolution. Hence I would like to propose that miRNAs are directly involved in maintenance of genomic integrity, and suggest a new role for them.

As we recently outlined in our latest study on the evolutionary origins of miRNAs, over a hundred miRNAs in the human genome were found as physically overlapping transposable elements (TEs) (Dahary *et al.* 2009). In fact, a large number of the newly discovered primate specific miRNAs were shown to originate from primate-specific and mammalian repeats, including *Alu* repeats (Lehnert *et al.* 2009), *MITEs* (Piriyapongsa and Jordan 2007) and others. Furthermore, miRNAs were previously suggested to target human and other mammalian TEs (Hakim *et al.* 2008; Lehnert *et al.* 2009) and the degree of retrotransposon activity in the human genome was found to be inversely correlated with the number of miRNAs predicted to target it (Hakim *et al.* 2008).

Given the above, I propose that miRNAs serve as guardians of genomic integrity by serving as global repressors of transposable elements. This idea is supported by

additional lines of evidence, particularly from other species, but also from the mammalian germline, as will be detailed below.

The RNAi pathway was implicated in silencing of repetitive DNA in many animal and plant species throughout the evolutionary tree, from the yeast *S. pombe* (Volpe *et al.* 2002), through nematodes (Tabara *et al.* 1999) to mammals (reviewed in (Slotkin and Martienssen 2007)). Knockout of *Dicer* in ES cells resulted in defective silencing of centromeric repeat sequences (Kanellopoulou *et al.* 2005). A newly identified class of *PIWI*-interacting siRNAs (later termed piRNAs) was found to be highly expressed in mouse oocytes and to mediate silencing of retrotransposons (Tam *et al.* 2008; Watanabe *et al.* 2008).

Expression of repeats and retrotransposons is widely associated with DNA damage in many species (Girard and Freeling 1999) (also reviewed in (Slotkin and Martienssen 2007)). Soper *et al*. have recently shown that de-repression of the transcription of transposable elements in mouse spermatocytes resulted in massive DNA damage (Soper *et al.* 2008).

Thus, multiple lines of evidence point to a role for the RNAi pathway as protector against retrotransposon expression and expansion, and accordingly also against the potential DNA damage induced by them, in multiple species as well as in the mammalian germline. This role resembles the ancestral role of RNAi that is common to all species having the RNAi machinery (reviewed in (Cerutti and Casas-Mollano 2006)). This role is not necessarily dependent on miRNAs genes. However, given the vast number of miRNAs that physically originate from TEs, it is intriguing to speculate that these also function in the repression of their "originators", not only in the germline but in other tissues as well.

Following the speculation that miRNAs are actively involved in guarding the stability of the genome through repression of TEs, one must address the question of how they exert this repression. As miRNAs were incriminated in both post-transcriptional gene silencing and in transcriptional gene silencing in other species (as elaborated

below), I would like to suggest that they also employ both strategies in achieving the global repression of TEs in their role in maintaining genomic integrity.

The RNAi machinery was implicated in transcriptional silencing of DNA primarily through Histone methylation in many species (reviewed in (Cerutti and Casas-Mollano 2006)). Involvement of the RNAi machinery in histone H3 methylation and epigenetic formation of heterochromatin was shown in several species (reviewed in (Grewal and Moazed 2003)) from plants (Zilberman *et al.* 2003) to *Drosophila* (Pal-Bhadra *et al.* 2002; Pal-Bhadra *et al.* 2004). The RNAi machinery was also implicated in promoting DNA methylation (reviewed in (Matzke *et al.* 2001)). The process of piRNA-mediated silencing of retrotransposon genes in the mouse germline was also shown to be mediated through *de novo* DNA methylation (Kuramochi-Miyagawa *et al.* 2008). Most importantly, in human cells, Morris *et al*. first showed that promoter methylation can be directed by introduction of a complementary siRNA, which served to direct the methylation towards a desired target sequence (Morris *et al.* 2004). This was later demonstrated to mediate long term silencing by DNA and histone methylation, and require protein components of the RNAi machinery (Hawkins *et al.* 2009).

In summary, and in light of the above evidence, I suggest that in normal tissues too, global repression of TEs can be mediated by TE-originated miRNAs, and may be driven not only through the well-known miRNA-mediated post-transcriptional silencing, but perhaps also through active involvement in epigenetic silencing of DNA: via DNA and histone methylation. I propose that just like in many other species and in the mouse germline, where there is an important role for the RNAi machinery in DNA and histone methylation that mediate TE repression, the same process might happen also in adult cells, and help to maintain the integrity of the genome in normal cells. When abolished, it may enable emerging tumor cells to increase their own genomic instability, thereby promoting tumorigenicity.

References

Adhikary, S. and M. Eilers (2005). "Transcriptional regulation and transformation by Myc proteins." Nat Rev Mol Cell Biol **6**(8): 635-45.

Alvarez-Garcia, I. and E. A. Miska (2005). "MicroRNA functions in animal development and human disease." Development **132**(21): 4653-62.

Ambros, V. and H. R. Horvitz (1984). "Heterochronic mutants of the nematode Caenorhabditis elegans." Science **226**(4673): 409-16.

Avraham, R., A. Sas-Chen, et al. (2009). "microRNAs Restrain Oncogenic Transcription Factors Induced by EGF." Submitted.

Aylon, Y., D. Michael, et al. (2006). "A positive feedback loop between the p53 and Lats2 tumor suppressors prevents tetraploidization." Genes Dev **20**(19): 2687-700.

Baek, D., J. Villen, et al. (2008). "The impact of microRNAs on protein output." Nature **455**(7209): 64-71.

Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-97.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-33.

Bernstein, E., S. Y. Kim, et al. (2003). "Dicer is essential for mouse development." Nat Genet **35**(3): 215-7.

Bommer, G. T., I. Gerin, et al. (2007). "p53-mediated activation of miRNA34 candidate tumor-suppressor genes." Curr Biol **17**(15): 1298-307.

Bonauer, A., G. Carmona, et al. (2009). "MicroRNA-92a Controls Angiogenesis and Functional Recovery of Ischemic Tissues in Mice." Science.

Borchert, G. M., W. Lanier, et al. (2006). "RNA polymerase III transcribes human microRNAs." Nat Struct Mol Biol **13**(12): 1097-101.

Borneman, A. R., J. A. Leigh-Bell, et al. (2006). "Target hub proteins serve as master regulators of development in yeast." Genes Dev **20**(4): 435-48.

Brennecke, J., A. Stark, et al. (2005). "Principles of microRNA-target recognition." PLoS Biol **3**(3): e85.

Brosh, R., R. Shalgi, et al. (2008). "p53-Repressed miRNAs are involved with E2F in a feed-forward loop promoting proliferation." Mol Syst Biol **4**: 229.

Brown, J. P., W. Wei, et al. (1997). "Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts." Science **277**(5327): 831-4.

Bueno, M. J., I. P. de Castro, et al. (2008). "Control of cell proliferation pathways by microRNAs." Cell Cycle **7**(20): 3143-8.

Bueno, M. J., I. Perez de Castro, et al. (2008). "Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression." Cancer Cell **13**(6): 496-506.

Bushati, N. and S. M. Cohen (2007). "microRNA functions." Annu Rev Cell Dev Biol **23**: 175-205.

Calin, G. A., C. D. Dumitru, et al. (2002). "Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia." Proc Natl Acad Sci U S A **99**(24): 15524-9.

Calin, G. A., C. Sevignani, et al. (2004). "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers." Proc Natl Acad Sci U S A **101**(9): 2999-3004.

Carleton, M., M. A. Cleary, et al. (2007). "MicroRNAs and cell cycle regulation." Cell Cycle **6**(17): 2127-32.

Cerutti, H. and J. A. Casas-Mollano (2006). "On the origin and functions of RNA-mediated silencing: from protists to man." Curr Genet **50**(2): 81-99.

Chang, T. C., E. A. Wentzel, et al. (2007). "Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis." Mol Cell **26**(5): 745-52.

Chang, T. C., D. Yu, et al. (2008). "Widespread microRNA repression by Myc contributes to tumorigenesis." Nat Genet **40**(1): 43-50.

Chaussepied, M. and D. Ginsberg (2005). "E2F and signal transduction pathways." Cell Cycle **4**(3): 392-6.

Chen, J. F., E. P. Murchison, et al. (2008). "Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure." Proc Natl Acad Sci U S A **105**(6): 2111-6.

Chi, S. W., J. B. Zang, et al. (2009). "Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps." Nature **460**(7254): 479-86.

Chivukula, R. R. and J. T. Mendell (2008). "Circular reasoning: microRNAs and cell-cycle control." Trends Biochem Sci **33**(10): 474-81.

Christoffersen, N. R., R. Shalgi, et al. (2009). "p53-independent upregulation of miR-34a during oncogene-induced senescence represses MYC." Cell Death Differ.

Cohen, E. E., H. Zhu, et al. (2009). "A feed-forward loop involving protein kinase Calpha and microRNAs regulates tumor cell cycle." Cancer Res **69**(1): 65-74.

Cullen, B. R. (2004). "Transcription and processing of human microRNA precursors." Mol Cell **16**(6): 861-5.

Dahary, D., R. Shalgi, et al. (2009). "A likely evolutionary origin of microRNAs in animals." Submitted.

Damiani, D., J. J. Alexander, et al. (2008). "Dicer inactivation leads to progressive functional and structural degeneration of the mouse retina." J Neurosci **28**(19): 4878-87.

Datta, J., H. Kutay, et al. (2008). "Methylation mediated silencing of MicroRNA-1 gene and its role in hepatocellular carcinogenesis." Cancer Res **68**(13): 5049-58.

de Alboran, I. M., R. C. O'Hagan, et al. (2001). "Analysis of C-MYC function in normal cells via conditional gene-targeted mutation." Immunity **14**(1): 45-55.

Dennis, G., Jr., B. T. Sherman, et al. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome Biol **4**(5): P3.

Di Micco, R., M. Fumagalli, et al. (2007). "Breaking news: high-speed race ends in arrest--how oncogenes induce senescence." Trends Cell Biol **17**(11): 529-36.

Dieci, G., G. Fiorino, et al. (2007). "The expanding RNA polymerase III transcriptome." Trends Genet **23**(12): 614-22.

Drayton, S., J. Rowe, et al. (2003). "Tumor suppressor p16INK4a determines sensitivity of human cells to transformation by cooperating cellular oncogenes." Cancer Cell **4**(4): 301-10.
el-Deiry, W. S., T. Tokino, et al. (1993). "WAF1, a potential mediator of p53 tumor suppression." Cell **75**(4): 817-25.
Enright, A. J., B. John, et al. (2003). "MicroRNA targets in Drosophila." Genome Biol **5**(1): R1.
Erson, A. E. and E. M. Petty (2008). "MicroRNAs in development and disease." Clin Genet **74**(4): 296-306.
Farh, K. K., A. Grimson, et al. (2005). "The widespread impact of mammalian MicroRNAs on mRNA repression and evolution." Science **310**(5755): 1817-21.
Filipowicz, W. (2005). "RNAi: the nuts and bolts of the RISC machine." Cell **122**(1): 17-20.
Frankel, L. B., N. R. Christoffersen, et al. (2008). "Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells." J Biol Chem **283**(2): 1026-33.
Gangaraju, V. K. and H. Lin (2009). "MicroRNAs: key regulators of stem cells." Nat Rev Mol Cell Biol **10**(2): 116-25.
Ginsberg, D. (2002). "E2F1 pathways to apoptosis." FEBS Lett **529**(1): 122-5.
Girard, L. and M. Freeling (1999). "Regulatory changes as a consequence of transposon insertion." Dev Genet **25**(4): 291-6.
Grewal, S. I. and D. Moazed (2003). "Heterochromatin and epigenetic control of gene expression." Science **301**(5634): 798-802.
Grimson, A., K. K. Farh, et al. (2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing." Mol Cell **27**(1): 91-105.
Grimson, A., M. Srivastava, et al. (2008). "Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals." Nature **455**(7217): 1193-7.
Hakim, S. T., M. Alsayari, et al. (2008). "A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses." Biochem Biophys Res Commun **369**(2): 357-62.
Hammond, S. M., S. Boettcher, et al. (2001). "Argonaute2, a link between genetic and biochemical analyses of RNAi." Science **293**(5532): 1146-50.
Han, J., Y. Lee, et al. (2004). "The Drosha-DGCR8 complex in primary microRNA processing." Genes Dev **18**(24): 3016-27.
Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.
Haupt, Y., R. Maya, et al. (1997). "Mdm2 promotes the rapid degradation of p53." Nature **387**(6630): 296-9.
Hawkins, P. G., S. Santoso, et al. (2009). "Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells." Nucleic Acids Res **37**(9): 2984-95.
He, L., X. He, et al. (2007). "A microRNA component of the p53 tumour suppressor network." Nature **447**(7148): 1130-4.
He, L., J. M. Thomson, et al. (2005). "A microRNA polycistron as a potential human oncogene." Nature **435**(7043): 828-33.

Heo, I., C. Joo, et al. (2008). "Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA." Mol Cell **32**(2): 276-84.
Ho, J. S., W. Ma, et al. (2005). "p53-Dependent transcriptional repression of c-myc is required for G1 cell cycle arrest." Mol Cell Biol **25**(17): 7423-31.
Hornstein, E., J. H. Mansfield, et al. (2005). "The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development." Nature **438**(7068): 671-4.
Hornstein, E. and N. Shomron (2006). "Canalization of development by microRNAs." Nat Genet **38 Suppl**: S20-4.
Houbaviy, H. B., M. F. Murray, et al. (2003). "Embryonic stem cell-specific MicroRNAs." Dev Cell **5**(2): 351-8.
Ishida, S., E. Huang, et al. (2001). "Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis." Mol Cell Biol **21**(14): 4684-99.
Ivanovska, I. and M. A. Cleary (2008). "Combinatorial microRNAs: working together to make a difference." Cell Cycle **7**(20): 3137-42.
Jeong, H., B. Tombor, et al. (2000). "The large-scale organization of metabolic networks." Nature **407**(6804): 651-4.
John, B., A. J. Enright, et al. (2004). "Human MicroRNA targets." PLoS Biol **2**(11): e363.
Johnson, S. M., H. Grosshans, et al. (2005). "RAS is regulated by the let-7 microRNA family." Cell **120**(5): 635-47.
Jovanovic, M. and M. O. Hengartner (2006). "miRNAs and apoptosis: RNAs to die for." Oncogene **25**(46): 6176-87.
Kanellopoulou, C., S. A. Muljo, et al. (2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing." Genes Dev **19**(4): 489-501.
Karolchik, D., R. Baertsch, et al. (2003). "The UCSC Genome Browser Database." Nucleic Acids Res **31**(1): 51-4.
Kent, O. A. and J. T. Mendell (2006). "A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes." Oncogene **25**(46): 6188-96.
Kim, V. N. (2005). "MicroRNA biogenesis: coordinated cropping and dicing." Nat Rev Mol Cell Biol **6**(5): 376-85.
Knoepfler, P. S. (2007). "Myc goes global: new tricks for an old oncogene." Cancer Res **67**(11): 5061-3.
Kong, Y. W., I. G. Cannell, et al. (2008). "The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene." Proc Natl Acad Sci U S A **105**(26): 8866-71.
Krek, A., D. Grun, et al. (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.
Krutzfeldt, J., N. Rajewsky, et al. (2005). "Silencing of microRNAs in vivo with 'antagomirs'." Nature **438**(7068): 685-9.
Kumamoto, K., E. A. Spillare, et al. (2008). "Nutlin-3a activates p53 to both down-regulate inhibitor of growth 2 and up-regulate mir-34a, mir-34b, and mir-34c expression, and induce senescence." Cancer Res **68**(9): 3193-203.

Kumar, M. S., J. Lu, et al. (2007). "Impaired microRNA processing enhances cellular transformation and tumorigenesis." Nat Genet **39**(5): 673-7.
Kuramochi-Miyagawa, S., T. Watanabe, et al. (2008). "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes." Genes Dev **22**(7): 908-17.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Laurent, L. C., J. Chen, et al. (2008). "Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence." Stem Cells **26**(6): 1506-16.
Lee, E. J., M. Baek, et al. (2008). "Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors." Rna **14**(1): 35-42.
Lee, R. C., R. L. Feinbaum, et al. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-54.
Lee, Y., C. Ahn, et al. (2003). "The nuclear RNase III Drosha initiates microRNA processing." Nature **425**(6956): 415-9.
Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." Embo J **23**(20): 4051-60.
Lehnert, S., P. Van Loo, et al. (2009). "Evidence for co-evolution between human microRNAs and Alu-repeats." PLoS One **4**(2): e4456.
Leone, G., J. DeGregori, et al. (1998). "E2F3 activity is regulated during the cell cycle and is required for the induction of S phase." Genes Dev **12**(14): 2120-30.
Lewis, B. P., C. B. Burge, et al. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." Cell **120**(1): 15-20.
Lewis, B. P., I. H. Shih, et al. (2003). "Prediction of mammalian microRNA targets." Cell **115**(7): 787-98.
Li, X., J. J. Cassidy, et al. (2009). "A microRNA imparts robustness against environmental fluctuation during development." Cell **137**(2): 273-82.
Li, Y., W. Tan, et al. (2009). "Role of the miR-106b-25 microRNA cluster in hepatocellular carcinoma." Cancer Sci.
Lim, L. P., N. C. Lau, et al. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-73.
Lim, L. P., N. C. Lau, et al. (2003). "The microRNAs of Caenorhabditis elegans." Genes Dev **17**(8): 991-1008.
Linhart, C., Y. Halperin, et al. (2008). "Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets." Genome Res **18**(7): 1180-9.
Liu, J., M. A. Carmell, et al. (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-41.
Lu, J., G. Getz, et al. (2005). "MicroRNA expression profiles classify human cancers." Nature **435**(7043): 834-8.
Lu, J., S. Guo, et al. (2008). "MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors." Dev Cell **14**(6): 843-53.

Lujambio, A., G. A. Calin, et al. (2008). "A microRNA DNA methylation signature for human cancer metastasis." Proc Natl Acad Sci U S A **105**(36): 13556-61.
Ma, Y., R. Croxton, et al. (2002). "Identification of novel E2F1-regulated genes by microarray." Arch Biochem Biophys **399**(2): 212-24.
Mangan, S. and U. Alon (2003). "Structure and function of the feed-forward loop network motif." Proc Natl Acad Sci U S A **100**(21): 11980-5.
Marais, G. (2003). "Biased gene conversion: implications for genome and sex evolution." Trends Genet **19**(6): 330-8.
Marson, A., S. S. Levine, et al. (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-33.
Martinez, N. J. and A. J. Walhout (2009). "The interplay between transcription factors and microRNAs in genome-scale regulatory networks." Bioessays **31**(4): 435-45.
Matzke, M., A. J. Matzke, et al. (2001). "RNA: guiding gene silencing." Science **293**(5532): 1080-3.
Meister, G., M. Landthaler, et al. (2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Mol Cell **15**(2): 185-97.
Mendell, J. T. (2008). "miRiad roles for the miR-17-92 cluster in development and disease." Cell **133**(2): 217-22.
Migliazza, A., F. Bosch, et al. (2001). "Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia." Blood **97**(7): 2098-104.
Migliazza, A., E. Cayanis, et al. (2000). "Molecular pathogenesis of B-cell chronic lymphocytic leukemia: analysis of 13q14 chromosomal deletions." Curr Top Microbiol Immunol **252**: 275-84.
Milo, R., S. Shen-Orr, et al. (2002). "Network motifs: simple building blocks of complex networks." Science **298**(5594): 824-7.
Milyavsky, M., Y. Tabach, et al. (2005). "Transcriptional programs following genetic alterations in p53, INK4A, and H-Ras genes along defined stages of malignant transformation." Cancer Res **65**(11): 4530-43.
Mooi, W. J. and D. S. Peeper (2006). "Oncogene-induced cell senescence--halting on the road to cancer." N Engl J Med **355**(10): 1037-46.
Morris, K. V., S. W. Chan, et al. (2004). "Small interfering RNA-induced transcriptional gene silencing in human cells." Science **305**(5688): 1289-92.
Mudhasani, R., Z. Zhu, et al. (2008). "Loss of miRNA biogenesis induces p19Arf-p53 signaling and senescence in primary cells." J Cell Biol **181**(7): 1055-63.
Muller, H., A. P. Bracken, et al. (2001). "E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis." Genes Dev **15**(3): 267-85.
Murchison, E. P., J. F. Partridge, et al. (2005). "Characterization of Dicer-deficient murine embryonic stem cells." Proc Natl Acad Sci U S A **102**(34): 12135-40.
Murchison, E. P., P. Stein, et al. (2007). "Critical roles for Dicer in the female germline." Genes Dev **21**(6): 682-93.
Murray-Zmijewski, F., E. A. Slee, et al. (2008). "A complex barcode underlies the heterogeneous response of p53 to stress." Nat Rev Mol Cell Biol **9**(9): 702-12.

Newman, M. A., J. M. Thomson, et al. (2008). "Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing." Rna **14**(8): 1539-49.
Nielsen, C. B., N. Shomron, et al. (2007). "Determinants of targeting by endogenous and exogenous microRNAs and siRNAs." Rna **13**(11): 1894-910.
Niwa, R. and F. J. Slack (2007). "The evolution of animal microRNA function." Curr Opin Genet Dev **17**(2): 145-50.
O'Donnell, K. A., E. A. Wentzel, et al. (2005). "c-Myc-regulated microRNAs modulate E2F1 expression." Nature **435**(7043): 839-43.
Olsen, P. H. and V. Ambros (1999). "The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation." Dev Biol **216**(2): 671-80.
Oren, M. (2003). "Decision making by p53: life, death and cancer." Cell Death Differ **10**(4): 431-42.
Pal-Bhadra, M., U. Bhadra, et al. (2002). "RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila." Mol Cell **9**(2): 315-27.
Pal-Bhadra, M., B. A. Leibovitch, et al. (2004). "Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery." Science **303**(5658): 669-72.
Parker, R. and U. Sheth (2007). "P bodies and the control of mRNA translation and degradation." Mol Cell **25**(5): 635-46.
Pilpel, Y., P. Sudarsanam, et al. (2001). "Identifying regulatory networks by combinatorial analysis of promoter elements." Nat Genet **29**(2): 153-9.
Piriyapongsa, J. and I. K. Jordan (2007). "A family of human microRNA genes from miniature inverted-repeat transposable elements." PLoS One **2**(2): e203.
Piriyapongsa, J., L. Marino-Ramirez, et al. (2007). "Origin and evolution of human microRNAs from transposable elements." Genetics **176**(2): 1323-37.
Polager, S. and D. Ginsberg (2008). "E2F - at the crossroads of life and death." Trends Cell Biol **18**(11): 528-35.
Polager, S., Y. Kalma, et al. (2002). "E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis." Oncogene **21**(3): 437-46.
Ragimov, N., A. Krauskopf, et al. (1993). "Wild-type but not mutant p53 can repress transcription initiation in vitro by interfering with the binding of basal transcription factors to the TATA motif." Oncogene **8**(5): 1183-93.
Rana, T. M. (2007). "Illuminating the silence: understanding the structure and function of small RNAs." Nat Rev Mol Cell Biol **8**(1): 23-36.
Raver-Shapira, N., E. Marciano, et al. (2007). "Transcriptional activation of miR-34a contributes to p53-mediated apoptosis." Mol Cell **26**(5): 731-43.
Re, A., D. Cora, et al. (2009). "Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human." Mol Biosyst **5**(8): 854-67.
Ren, B., H. Cam, et al. (2002). "E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints." Genes Dev **16**(2): 245-56.
Rosenfeld, N., R. Aharonov, et al. (2008). "MicroRNAs accurately identify cancer tissue origin." Nat Biotechnol **26**(4): 462-9.

Sachdeva, M., S. Zhu, et al. (2009). "p53 represses c-Myc through induction of the tumor suppressor miR-145." Proc Natl Acad Sci U S A **106**(9): 3207-12.
Saini, H. K., S. Griffiths-Jones, et al. (2007). "Genomic analysis of human microRNA transcripts." Proc Natl Acad Sci U S A **104**(45): 17719-24.
Sandberg, R., J. R. Neilson, et al. (2008). "Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites." Science **320**(5883): 1643-7.
Sandelin, A., P. Carninci, et al. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." Nat Rev Genet **8**(6): 424-36.
Sasaki, T., A. Shiohama, et al. (2003). "Identification of eight members of the Argonaute family in the human genome small star, filled." Genomics **82**(3): 323-30.
Sears, R., K. Ohtani, et al. (1997). "Identification of positively and negatively acting elements regulating expression of the E2F2 gene in response to cell growth signals." Mol Cell Biol **17**(9): 5227-35.
Segal, E., M. Shapira, et al. (2003). "Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data." Nat Genet **34**(2): 166-76.
Selbach, M., B. Schwanhausser, et al. (2008). "Widespread changes in protein synthesis induced by microRNAs." Nature **455**(7209): 58-63.
Sempere, L. F., C. N. Cole, et al. (2006). "The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint." J Exp Zoolog B Mol Dev Evol **306**(6): 575-88.
Seoane, J., H. V. Le, et al. (2002). "Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage." Nature **419**(6908): 729-34.
Serrano, M., A. W. Lin, et al. (1997). "Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a." Cell **88**(5): 593-602.
Sethupathy, P., B. Corda, et al. (2006). "TarBase: A comprehensive database of experimentally supported animal microRNA targets." Rna **12**(2): 192-7.
Shalgi, R. (2009). "Novel function of miRNAs: maintenance of genomic integrity in mammals through repression of transposable elements." In Preparation.
Shalgi, R., R. Brosh, et al. (2009). "Coupling transcriptional and post-transcriptional miRNA regulation in the control of cell fate." Submitted.
Shalgi, R., D. Lieber, et al. (2007). "Global and Local Architecture of the Mammalian microRNA-Transcription Factor Regulatory Network." PLoS Comput Biol **3**(7): e131.
Shen-Orr, S. S., R. Milo, et al. (2002). "Network motifs in the transcriptional regulation network of Escherichia coli." Nat Genet **31**(1): 64-8.
Sinha, S., A. S. Adler, et al. (2008). "Systematic functional characterization of cis-regulatory motifs in human core promoters." Genome Res **18**(3): 477-88.
Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-85.
Smibert, P. and E. C. Lai (2008). "Lessons from microRNA mutants in worms, flies and mice." Cell Cycle **7**(16): 2500-8.

Song, J. J., J. Liu, et al. (2003). "The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes." Nat Struct Biol **10**(12): 1026-32.

Soper, S. F., G. W. van der Heijden, et al. (2008). "Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis." Dev Cell **15**(2): 285-97.

Sorlie, T., Y. Wang, et al. (2006). "Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms." BMC Genomics **7**: 127.

Stanelle, J., T. Stiewe, et al. (2002). "Gene expression changes in response to E2F1 activation." Nucleic Acids Res **30**(8): 1859-67.

Stark, A., J. Brennecke, et al. (2005). "Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution." Cell **123**(6): 1133-46.

Stark, A., J. Brennecke, et al. (2003). "Identification of Drosophila MicroRNA targets." PLoS Biol **1**(3): E60.

Sun, F., H. Fu, et al. (2008). "Downregulation of CCND1 and CDK6 by miR-34a induces cell cycle arrest." FEBS Lett **582**(10): 1564-8.

Suzuki, H. I., K. Yamagata, et al. (2009). "Modulation of microRNA processing by p53." Nature **460**(7254): 529-33.

Sylvestre, Y., V. De Guire, et al. (2007). "An E2F/miR-20a autoregulatory feedback loop." J Biol Chem **282**(4): 2135-43.

Tabara, H., M. Sarkissian, et al. (1999). "The rde-1 gene, RNA interference, and transposon silencing in C. elegans." Cell **99**(2): 123-32.

Tam, O. H., A. A. Aravin, et al. (2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes." Nature **453**(7194): 534-8.

Tanzer, A. and P. F. Stadler (2004). "Molecular evolution of a microRNA cluster." J Mol Biol **339**(2): 327-35.

Tarasov, V., P. Jung, et al. (2007). "Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest." Cell Cycle **6**(13): 1586-93.

Tazawa, H., N. Tsuchiya, et al. (2007). "Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells." Proc Natl Acad Sci U S A **104**(39): 15472-7.

Thai, T. H., D. P. Calado, et al. (2007). "Regulation of the germinal center response by microRNA-155." Science **316**(5824): 604-8.

Trumpp, A., Y. Refaeli, et al. (2001). "c-Myc regulates mammalian body size by controlling cell number but not cell size." Nature **414**(6865): 768-73.

Tsang, J., J. Zhu, et al. (2007). "MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals." Mol Cell **26**(5): 753-67.

van den Berg, A., J. Mols, et al. (2008). "RISC-target interaction: cleavage and translational suppression." Biochim Biophys Acta **1779**(11): 668-77.

Ventura, A., A. G. Young, et al. (2008). "Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters." Cell **132**(5): 875-86.

Viswanathan, S. R., G. Q. Daley, et al. (2008). "Selective blockade of microRNA processing by Lin28." Science **320**(5872): 97-100.
Viswanathan, S. R., J. T. Powers, et al. (2009). "Lin28 promotes transformation and is associated with advanced human malignancies." Nat Genet.
Volinia, S., G. A. Calin, et al. (2006). "A microRNA expression signature of human solid tumors defines cancer gene targets." Proc Natl Acad Sci U S A **103**(7): 2257-61.
Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-7.
Voorhoeve, P. M., C. le Sage, et al. (2006). "A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors." Cell **124**(6): 1169-81.
Vousden, K. H. and C. Prives (2009). "Blinded by the Light: The Growing Complexity of p53." Cell **137**(3): 413-31.
Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-43.
Weinmann, A. S., S. M. Bartley, et al. (2001). "Use of chromatin immunoprecipitation to clone novel E2F target promoters." Mol Cell Biol **21**(20): 6820-32.
Wienholds, E., M. J. Koudijs, et al. (2003). "The microRNA-producing enzyme Dicer1 is essential for zebrafish development." Nat Genet **35**(3): 217-8.
Wightman, B., I. Ha, et al. (1993). "Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans." Cell **75**(5): 855-62.
Winter, J., S. Jung, et al. (2009). "Many roads to maturity: microRNA biogenesis pathways and their regulation." Nat Cell Biol **11**(3): 228-34.
Woods, K., J. M. Thomson, et al. (2007). "Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors." J Biol Chem **282**(4): 2130-4.
Xi, Y., R. Shalgi, et al. (2006). "Differentially regulated micro-RNAs and actively translated messenger RNA transcripts by tumor suppressor p53 in colon cancer." Clin Cancer Res **12**(7 Pt 1): 2014-24.
Zhang, B., X. Pan, et al. (2007). "microRNAs as oncogenes and tumor suppressors." Dev Biol **302**(1): 1-12.
Zhang, L., S. Volinia, et al. (2008). "Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer." Proc Natl Acad Sci U S A **105**(19): 7004-9.
Zhao, Y., E. Samal, et al. (2005). "Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis." Nature **436**(7048): 214-20.
Zhou, Y., J. Ferguson, et al. (2007). "Inter- and intra-combinatorial regulation by transcription factors and microRNAs." BMC Genomics **8**: 396.
Zhu, J., D. Woods, et al. (1998). "Senescence of human fibroblasts induced by oncogenic Raf." Genes Dev **12**(19): 2997-3007.
Zilberman, D., X. Cao, et al. (2003). "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." Science **299**(5607): 716-9.

Appendix

Differentially Regulated Micro-RNAs and Actively Translated Messenger RNA Transcripts by Tumor Suppressor p53 in Colon Cancer

See attached manuscript (Xi *et al.* 2006).

Acknowledgments

I would like to take this opportunity to thank the significant people who have influenced and helped me the most during this period of my PhD.

First and foremost – to my two advisors: Tzachi Pilpel and Moshe Oren, the most amazing two advisors I could ever have chosen.

Tzachi – I would like to thank you for teaching me how to "make science", for guiding me through my own path. For taking me through my baby steps and holding my hand until I learned to walk my own scientific path. For teaching me what systems biology really is, and how bioinformatic research is properly done: asking real biological questions. I have known you for almost seven years, back in the days of the Meyer basement, and feel we have come a long way. I would like to thank you for teaching me how to ask questions, and to pursue my own ideas and ambitions. I thank you for helping me stand up for my ideas, not give up on them, and sticking with me through tough situations, and for being there at all times, day or night. You taught me how to think as a scientist, work as a scientist, to be excited from the science – how to be a scientist. And for that I am grateful.

Moshe – I would like to thank you for teaching me what is true molecular biology, for teaching me to think deeply, and widely, both in and outside the box. I came to your office 4.5 years ago with a huge enthusiasm, and very little knowledge of molecular biology and techniques, and learned so much, without ever hearing that my ideas are too wild or not good. I enjoyed and learned from every single conversation, and thanks to you now I have the deep biological understanding that allows me to ask biological questions on my own. Thank you for teaching me how fascinating molecular biology can be, and to love it so much. Now I know that classical biology can (and must) go hand in hand with systems biology, and can be as fun and interesting. For all that I am grateful.

Tzachi and Moshe – to both of you I would like to thank for always encouraging me to follow my own path and pursue my own ideas, for never discouraging my ambitions, and

always pushing me forward. I feel today that with the combination of you both as mentors I got all that I could possibly want and need in a PhD, and much more. You two helped me stand on my own two scientific feet, and thanks to you both I now come out to the world excited, motivated, and confident.

I would also like to thank my many collaborators, and specifically Ran Brosh and Dvir Dahary, both of you have broadened my horizons in different ways. Ran, collaborating with you has been a wonderful experience, I have learned a lot and had many laughs along the way. I am very proud and happy with our end result, and with the way we achieved it together, and for that I thank you. Dvir, thank you for many fun, and inspiring discussions, for teaching me a lot about evolution, and for bringing your unique perspective which many times helped me to improve my own.

Many thanks to the members of the Oren lab and the Pilpel lab, who provided a lot of help, technical support, emotional support, experimental support, and a fun environment to work in. Thanks to all the members, past and present, and most of all to the friends who accompany me from day one in the lab: Yael, Sylvie, Neta, Efrat and the rest of the Oren lab, Orna, Ilya, Dvir and the rest of the Pilpel lab.

And last but not least, I would like to thank my family. To my mom and dad (and baby brother) for always believing in me, motivating me and pushing me to pursue my dreams and desires, for being there and supporting me during the tough times, and for ever being so proud. I am where I am today because of you, and I thank you for it and love you a lot. And to Alon, my light, shiny optimistic other half, for the endless love, for always being there, for being part of the whole process, and for your support through every step of the way, and the willingness to join me to the other side of the world, without which I would never have gotten this far.

Global and Local Architecture of the Mammalian microRNA–Transcription Factor Regulatory Network

Reut Shalgi[1,2], Daniel Lieber[1], Moshe Oren[2], Yitzhak Pilpel[1*]

1 Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel, **2** Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel

microRNAs (miRs) are small RNAs that regulate gene expression at the posttranscriptional level. It is anticipated that, in combination with transcription factors (TFs), they span a regulatory network that controls thousands of mammalian genes. Here we set out to uncover local and global architectural features of the mammalian miR regulatory network. Using evolutionarily conserved potential binding sites of miRs in human targets, and conserved binding sites of TFs in promoters, we uncovered two regulation networks. The first depicts combinatorial interactions between pairs of miRs with many shared targets. The network reveals several levels of hierarchy, whereby a few miRs interact with many other lowly connected miR partners. We revealed hundreds of "target hubs" genes, each potentially subject to massive regulation by dozens of miRs. Interestingly, many of these target hub genes are transcription regulators and they are often related to various developmental processes. The second network consists of miR–TF pairs that coregulate large sets of common targets. We discovered that the network consists of several recurring motifs. Most notably, in a significant fraction of the miR–TF coregulators the TF appears to regulate the miR, or to be regulated by the miR, forming a diversity of feed-forward loops. Together these findings provide new insights on the architecture of the combined transcriptional–post transcriptional regulatory network.

Citation: Shalgi R, Lieber D, Oren M, Pilpel Y (2007) Global and local architecture of the mammalian microRNA–transcription factor regulatory network. PLoS Comput Biol 3(7): e131. doi:10.1371/journal.pcbi.0030131

Introduction

microRNAs (miRs) are short RNAs that post transcriptionally regulate messenger RNAs. Two main mechanisms for such effects are degradation of the target mRNA, and inhibition of its translation [1]. In recent years considerable progress within multiple genomes was obtained in the experimental identification of genes encoding for miRs [2–4], and in tools for the identification of target genes of miRs, based on miR sequences and the sequence of the targets' 3′ untranslated regions (UTRs) [5–11]. Compared with the regulation of transcription, the study of the regulatory networks spanned by miRs is only at its beginning. When it comes to transcriptional regulation, a lot is known about the main players and the interactions between them. Transcription factors (TFs) are well-characterized [12], and promoter binding motifs are available in a diversity of species [13]. The combinatorial interactions between TFs have been explored [14,15] as well as the global level properties of the transcription regulatory network [16]. In addition, the local structures of the network have been intensively investigated. It was found in several species that the transcription regulatory network may be decomposed into elementary building blocks, or network motifs, that recur in the network more than expected by chance, and that these motifs likely perform local "computations," such as the detection of signal persistency or the coordinated gradual activation of a set of genes [17–20].

When it comes to posttranscriptional regulation, and in particular to the miR world, most of the parallel knowledge is lacking. While we do know about many miRs in multiple genomes [1], their targets are predicted with relatively limited accuracy [21]. Even more obvious is the lack of knowledge about the structure of the miR regulatory network, and about the potential interface between this network and the transcriptional one. In similarity to TFs, miRs are expected to work in combinations on their target genes [7]. The target specificity-determining site of the miRs is often short (seven to eight nucleotides [9]), hence some genes that contain a match to a single miR in their 3′ UTRs may represent false positive assignments. Thus, combinatorial interactions among the miRs are probably necessary to specify more precisely the set of affected targets of each miR. As in the realm of transcription regulators [14], combinatorics may also have the advantage of allowing multiple sources of information, each represented by a single miR, to be integrated into the regulation of individual transcripts.

Since TFs regulate mRNA production, and miRs regulate transcript stability and its translation, an attractive possibility is that miRs and TFs cooperate in regulating shared target genes. This possibility is appealing since a gene that is regulated through multiple mechanisms may be tuned at a level of precision that is higher than what may be obtained by

Editor: Gary Stormo, Washington University, United States of America

Received December 27, 2006; **Accepted** May 22, 2007; **Published** July 13, 2007

Abbreviations: FDR, false discovery rate; FFL, feed-forward loop; miR, microRNA; PSSM, position specific scoring matrices; TF, transcription factor; TFBS, TF binding sites; TSS, transcription start site; UTR, untranslated region; GO, Gene Ontology

* To whom correspondence should be addressed. E-mail: Pilpel@weizmann.ac.il

Author Summary

It is becoming increasingly appreciated that a new type of gene which does not code for proteins, the regulatory RNAs, constitutes a considerable portion of mammalian genomes, and these genes serve as key players in the regulatory network of living cells. Among these regulatory RNAs are the microRNAs (miRs), small RNAs that mediate posttranscriptional gene silencing through inhibition of protein production or degradation of mRNAs. So far little is known about the extent of regulation by miRs, and their potential cooperation with other regulatory layers in the network. We investigated the potential crosstalk between the miR-mediated posttranscription layer, and the transcriptional regulation layer, whose dominant players, the transcription factors (TFs), regulate the production of protein-coding mRNAs. We found that the extent of miR regulation varies extensively among different genes, some of which, especially those who serve as regulators themselves, are subject to enhanced miR silencing. Further, we identified thousands of genes that are potentially subjected to coordinated regulation by multiple miRs and by specific combinations of TFs and miRs. The regulatory network, comprising transcriptional and posttranscriptional regulation, manifests several recurring architectures, one of which consists of a TF and a miR that together regulate a large set of common genes, and that also appear to regulate one another. Altogether this work provides new insights into the logic and evolution of a new regulatory layer of the mammalian genome, and its effect on other regulatory networks in the cell.

either mechanism alone. In addition, as with any other regulatory agent in cells, the question "what regulates the regulator" is of prime importance, as it may allow the exposure of multiple levels of hierarchies and their design within a control network. It is thus crucial to understand whether TFs and miRs collaborate in gene regulation, and also to characterize regulatory interactions that miRs and TFs may exert on each other. In similarity to the transcription network, local network motifs might exist which may also consist of miRs. One attractive role for such motifs has been suggested in a developmental context—to canalize "noise" in gene expression [22]. However, actual realization of such motifs remains to be explored.

Here we report extensive combinatorial interactions among miRs and between miRs and TFs. We found hundreds of miRs target hubs—genes regulated by dozens of miRs—which are involved in a diversity of developmental processes and in transcription regulation. The miR–TF regulatory network features several motifs in which TF and miR partners that are suggested to regulate multiple target genes often exert regulation on one another.

Results

Connectivity Distributions in the miR–Gene Network

We used two datasets of miRs and their predicted target genes: TargetScan [8,9] and PicTar [7]. The miRs used in this analysis are characterized by being evolutionarily conserved, and, in addition, their targets were defined based on conservation in orthologous genes in four species (human, mouse, rat, and dog). This evolutionary conservation criterion was assumed to constitute a good filter for false positive assignments of miRs to genes [9,23]. Yet, it must be emphasized that the accuracy of such assignments is still limited [21] (see "noise tolerance analysis" in Materials and Methods). Altogether we analyzed 8,672 and 9,152 human (RefSeq) genes in the TargetScan and PicTar datasets, respectively, that have at least one predicted miR binding site in their 3′ UTR, and a total of 138 miRs and 178 miRs in the respective datasets.

We constructed a matrix whose rows are genes and columns are miRs, in which the ij-th element is "1" if gene i contains a predicted binding site for miR j in its 3′ UTR, and "0" otherwise. We created one such matrix for each of the two miR target prediction datasets. For the sake of clarity, from here on we will say interchangeably that "a miR targets a gene" or that "a gene contains in its 3′ UTR a predicted binding site for a miR." We first characterized the matrix by the distribution of degree connectivity of each gene and of each miR. Figure 1A shows the distribution of the number of miRs assigned per gene, while Figure 1B shows the distribution of number of genes assigned to each miR. We compared each distribution with a set of distributions, each derived by randomization of the original matrix according to two alternative null models. Along with the distribution of number of miRs per gene (Figure 1A), we also plotted 100 distributions obtained after randomizing each of the columns in the matrix. In this randomization we preserved the number of genes per miR, yet assigned genes at random to each miR. The distributions obtained after the randomization differed markedly from the original distribution, both in terms of width and shape. While in the randomized distributions genes rarely have more than ten different miRs in their 3′ UTR, in the original distribution there are hundreds of genes subjected to extensive predicted miR regulation. In Figure 1B we also show the distribution of number of genes per miR. Along with it is shown a set of distributions obtained by randomizing each of the rows in the matrix, namely by randomly assigning miRs to each gene, preserving the real number of miRs predicted to target each gene, as in the original matrix. Here, too, the randomized distributions differed from the original one both in shape and width; the original data contains multiple miRs which appear to target more than 400 genes, significantly higher than the number that would be obtained by merely preserving the statistics of number of miR sites in genes UTRs. These observations lead us to highlight some special properties that seem to be unique to the miR regulatory network.

Target Hubs—Genes with Extensive miR Regulation

The distribution of number of miRs regulating each target gene (Figure 1A) has a long right tail in contrast to the distributions in the randomized matrices that looked Gaussian (as befits a sum of independent random variables). We thus focused on the genes in that tail of the distribution (which are targeted by more than 15 miRs and 20 miRs in the TargetScan and PicTar datasets, respectively; see Materials and Methods for further details and cutoff justification). We named these genes target hubs following a recent definition of genes regulated by multiple TFs in yeast [24]. There are 470 such genes in the TargetScan dataset. We made similar observations with the PicTar dataset and identified 834 target hubs (see Figure S1)—the set of target hubs based on the TargetScan dataset has an 81% overlap with the target hubs defined by PicTar dataset.

Inspecting the target hubs genes' annotations (using Gene Ontology, GO), we found that they are highly enriched for

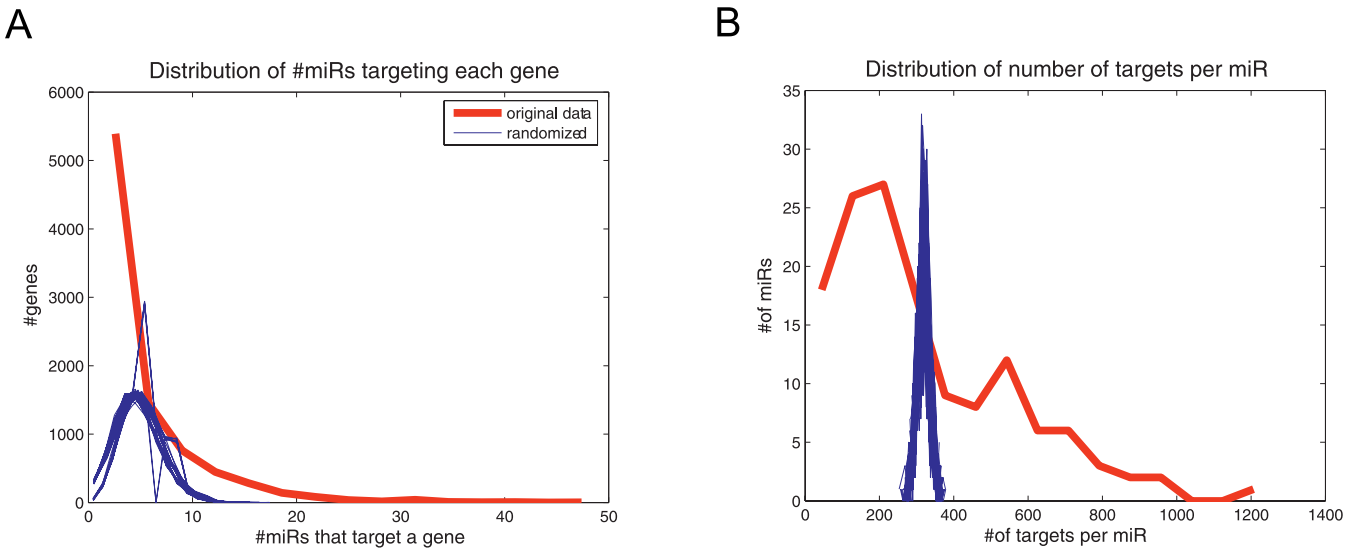


Figure 1. miRs and Target Genes in the TargetScan Dataset

(A) Distribution of the number of different miRs regulating each target gene in the TargetScan dataset. The thick red line represents the distribution in the original datasets, while each of the thin blue lines represents the distribution in one of the column-randomized matrices. The matrix contains only genes with at least one predicted site in their 3′ UTR. In each randomization, we shuffled the assignment of miRs to their targets, keeping constant the number of targets per miR.

(B) Distribution of number of targets per miR in the TargetScan dataset. In the thick red line we depicted the original distribution, while each blue thin line represents the distribution in one of the 100 row-randomized matrices, which preserve the distribution of number of miRs targeting each gene.

doi:10.1371/journal.pcbi.0030131.g001

developmental processes, specifically for muscle development and nervous system development, as well as for TFs and transcription regulators (see Table 1 for enrichment statistics). Among the transcription regulators in the set of target hubs are included RUNX1, E2F-3, N-MYC, and SP3. Another very intriguing fact is that the *Ago1* gene, one of the key components of the human RISC (RNAi induced silencing complex), is also a target hub, as in the dataset it appears to be potentially regulated by multiple miRs.

We suspected, however, that the fact that target hubs host many miR binding sites may result from potentially longer 3′ UTRs [23]. Although we found that target hubs have a distribution of 3′ UTR lengths that is significantly longer than that of the rest of the genes in the current analysis (p-value $= 4 \times 10^{-85}$ and p-value $= 3 \times 10^{-101}$ for TargetScan and PicTar target hubs, respectively, using the Kolmogorov-Smirnov test), we still realized that many of them have relatively short 3′ UTRs (Figure S2A and S2B). To test whether the high number of miR binding sites in the target hubs is a simple reflection of their 3′ UTR lengths, we performed a randomization test, in which we sampled 100 times random gene sets from the entire dataset with the same or very similar length distributions as that of the target hubs (see Materials and Methods). We found that such gene sets always have a significantly lower average number of miR sites per gene compared with the target hubs (see Figure S3A). We further calculated the density of different miRs in the 3′ UTRs [23]. Density was defined as number of different miRs targeting a gene divided by 3′ UTR length. Remarkably, we found that the miR density in the target hubs is significantly higher than in the rest of the genes in the dataset (p-value $= 2 \times 10^{-85}$ and p-value $= 6 \times 10^{-124}$ for the TargetScan and PicTar target hubs, respectively, using the Kolmogorov-Smirnov test; the means are 2.84 and 1.80 times higher in the TargetScan and PicTar dataset means, respectively; see Figure 2 and Figure S2C for the entire distributions). We concluded that target hubs are rich in binding sites for different miRs to an extent that cannot be explained solely by their 3′ UTRs lengths.

Realizing that density of miR binding sites may be an important property by itself, we also used an alternative definition for target hubs—genes with particularly high density of miRs in their 3′ UTRs. We collected the genes in the top 85th percentile of the miR binding site density spectrum, then we performed a similar GO enrichment analysis to see whether particular functionalities were enriched among the genes with a high density of miR binding sites. Reassuringly, most of the functionalities that were enriched among the set of target hubs defined by number of differnet miRs were also significant in the set of high density target hubs (see Table 1). Moreover, we found that genes that were target hubs according to only one of the two definitions (i.e., genes that are not in the overlap of the two sets) were still significantly enriched for functionalities such as transcription regulator activity and development (unpublished data).

A Combinatorial Network of miR Interactions

Combinatorial interactions are a fundamental property of the transcription networks [25]. It may be anticipated that, similarly to TFs, miRs may work in combinations. One way to predict pairs of coregulating miRs is to ask which pairs show a high rate of co-occurrence in the same target genes' 3′ UTRs. A common statistical test in the field, previously used in the context of promoter motifs and TF binding site [26–28], is the cumulative hypergeometric statistic. According to this model, given the rate of occurrence of each of the regulators alone, and the total number of genes in the analysis, a p-value is computed on the size of the set of genes that are shared between the two regulators. The main assumption of this

Table 1. TargetScan Target Hubs GO Functional Enrichment

GO Annotation	Target Hubs with High Number of miRs		High Density Target Hubs	
	Enrichment *p*-Value	Number of Target Hubs with Annotation	Enrichment *p*-Value	Number of Target Hubs with Annotation
Development	4.09E−07	48	3.74E−04	81
DNA binding	8.59E−09	61	1.31E−04	102
Muscle development	3.50E−05	8	NS	
Negative regulation of transcription, DNA-dependent	8.12E−05	8	NS	
Nervous system development	2.34E−10	25	4.39E−04	27
Nucleus	4.80E−06	83	4.07E−05	171
Protein binding	3.39E−06	94	6.90E−06	200
Regulation of transcription	1.49E−06	58	1.24E−06	117
Regulation of transcription from RNA polymerase II promoter	4.30E−03	9	2.04E−07	25
Regulation of transcription, DNA-dependent	3.13E−07	57	1.69E−07	114
RNA polymerase II TF activity	2.26E−06	13	2.36E−10	27
System development	2.77E−10	25	4.96E−04	27
Transcription	2.38E−06	59	9.78E−07	121
Transcription coactivator activity	1.15E−02	6	7.57E−07	18
Transcription cofactor activity	3.60E−03	9	1.62E−06	23
TF activity	7.02E−09	37	1.37E−05	57
TF binding	2.32E−03	11	1.67E−05	25
Transcription from RNA polymerase II promoter	2.88E−08	22	6.20E−09	38
Transcription regulator activity	1.82E−09	47	4.34E−08	83
Transcriptional activator activity	1.93E−03	9	1.57E−05	20

Target hubs were defined by two alternative definitions: target hubs with high number of miR binding site (more than 15 in the case of TargetScan and more than 20 in the case of PicTar), or as high density target hubs (genes with high density of miR binding sites in their 3′ UTRs). We used the standard method of hypergeometric *p*-value to test for functionally enriched GO annotations in each gene set. The results were corrected for multiple hypotheses and annotations were considered significantly enriched if they passed FDR of 0.05. We present here the union of significant annotations for the high density target hubs and the high miR number target hubs.

doi:10.1371/journal.pcbi.0030131.t001

model, that assignment of a gene to the first regulator is independent of the assignment to the second one, is likely fulfilled in the context of fixed-length promoters. Yet when it comes to 3′ UTRs of varying length, the assumption does not hold anymore. Some genes, e.g., those with long 3′ UTRs, have a higher chance to contain predicted binding sites for miRs, hence a *p*-value calculated based on the hypergeometric model may overestimate the significance of the co-occurrence rate.

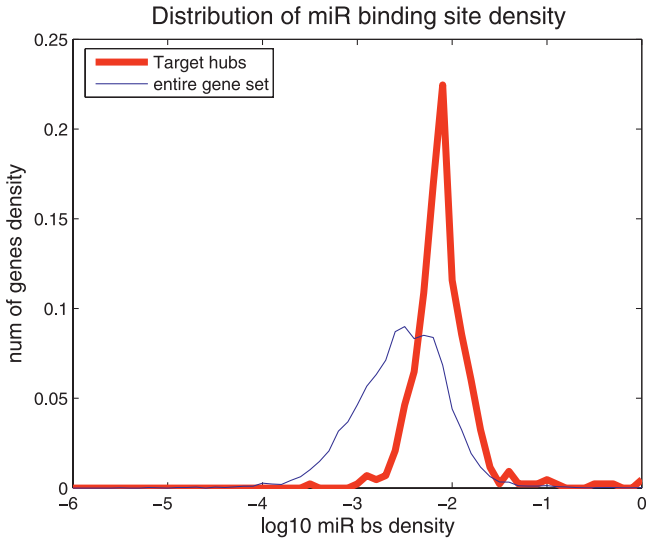


Figure 2. Distribution of the density of miRs in the 3′ UTRs of target hubs (thick red line) and all the genes (thin blue line) in the TargetScan dataset (all genes included in this figures have at least one miR site predicted in their 3′ UTR). The log10 densities were binned into bins of 0.1, and relative frequencies were plotted. Same analysis for the PicTar dataset is in Figure S2.

doi:10.1371/journal.pcbi.0030131.g002

We have thus devised an alternative, randomization-based test for identifying significantly co-occurring miR pairs. The model was designed such that it will capture the underlying distributions in Figure 1A and 1B, and test whether a given pair of miRs co-occurs at a higher rate, considering the above distributions as a background. For each pair of miRs, i and j, with their set of targets, Targets(i) and Targets(j), respectively, we calculated the "Meet/Min" score [29,30] defined in the present case as:

$$\frac{|\mathrm{Targets}(i) \cap \mathrm{Targets}(j)|}{\min(|\mathrm{Targets}(i)|, |\mathrm{Targets}(j)|)}$$

namely, the size of the set of genes that contain sites for the two miRs together, divided by the smaller of the two sets of targets (we filtered from the calculation for each i,j pair, 3′ UTRs in which the sites for i and j are physically overlapping to avoid overestimation of significance of miR pairs with an overlapping or similar seed, see Materials and Methods for details). Yet this score is not a statistic, i.e., it lacks an estimate of the probability to obtain such score (or better) by chance given an appropriate null model. Following previous works [20], we used a null model that preserves for each gene the number miRs assigned to it, and for each miR the number of genes assigned to it in the input data. We generated 1,000 randomized matrices according to this null model. In each such matrix we randomized the original matrix in 100,000

Table 2. Top 20 Most Significant Pairs of Coregulating miRNAs in the TargetScan Network

First miR	Number of Targets	Second miR	Number of Targets	Number of Common Genes	Meet/Min Score	Co-Occurrence *p*-Value
miR-133	317	miR-423	9	4	0.44	<10E−3
miR-147	8	miR-219	210	3	0.38	<10E−3
miR-146	103	miR-423	9	3	0.33	<10E−3
miR-30	737	miR-365	129	41	0.32	<10E−3
miR-362	94	miR-495	503	29	0.31	<10E−3
miR-23	514	miR-362	94	28	0.30	<10E−3
miR-185	135	miR-20	608	40	0.30	<10E−3
miR-181	578	miR-362	94	27	0.29	<10E−3
miR-29	567	miR-362	94	26	0.28	<10E−3
miR-221	211	miR-495	503	58	0.27	<10E−3
miR-25	474	miR-342	126	34	0.27	<10E−3
miR-200b	500	miR-382	88	23	0.26	<10E−3
miR-186	419	miR-362	94	24	0.26	<10E−3
miR-362	94	miR-93	369	24	0.26	<10E−3
miR-362	94	miR-369–3p	361	23	0.24	<10E−3
miR-129	197	miR-381	547	47	0.24	<10E−3
miR-218	454	miR-409–5p	93	22	0.24	<10E−3
miR-19	563	miR-329	182	43	0.24	<10E−3
miR-369–3p	361	miR-495	503	85	0.24	<10E−3

The table depicts the number of targets each miR has in the specific database, and the number of targets which contain sites for both miRs. We note that in each pairing, we filtered out genes where binding sites for the two miRs physically overlapped, so this *p*-value is not biased by miRs with highly similar seeds. For this reason, the number of target genes may be slightly different for the same miR in two different pairings. For the full list of significantly co-occurring miR pairs in the TargetScan and PicTar datasets, see Tables S1 and S2.
doi:10.1371/journal.pcbi.0030131.t002

steps, using an edge-swapping algorithm [20]. For each such randomized matrix we computed again the Meet/Min score for all pairs of miRs. The co-occurrence *p*-value for a pair of miRs was computed according to the pair's Meet/Min score and the population of 1,000 Meet/Min scores obtained for that same pair in each of the 1,000 edge-swapped matrices. The *p*-value for the pair is defined as the fraction of the 1,000 randomized matrices in which the Meet/Min score of that pair is greater than or equal to the Meet/Min score of the pair in the original matrix.

In addition to calculating a score of co-occurrence, we also calculated, using the same formalism, a score that captures the tendency of every two miRs to avoid residing within shared 3′ UTRs. We will regard a pair of miRs that co-occur in the original matrix significantly less frequently than in the edge-swapped matrices as avoiding each other. Given the Meet/Min score of co-occurrence for a pair of miRs, and the Meet/Min scores obtained for that pair in the 1,000 edge-swapped matrices, we calculated the fraction of randomized scores that were lower than or equal to that obtained in the original matrix for that pair, as the avoidance *p*-value of a miR pair.

In both cases of co-occurrence and avoidance, we used the false discovery rate (FDR) to control for the testing of multiple hypotheses. In the case of co-occurring miR pairs, using a restrictive FDR threshold (q-value = 0.05), we obtained 107 pairs with a significant *p*-value in the TargetScan dataset, and 199 pairs in the PicTar dataset (interestingly, the ratio between the number of interactions in the two datasets ($\sim$0.54) is very close to the ratio expected based on the square of relative number of miRs in each dataset ($\sim$0.6)). We created a combinatorial network based on the significant co-occurring miR pairs. The top miR pairs are given in Table 2 and are also depicted in Figures 3A and S4A. The full list of significant pairs is provided in Tables S1 and S2. This combinatorial network consists of several levels of hierarchy. At the top (Figure 3A) are a handful of miRs that interact with a relatively large number of miR partners, while at the bottom are "end-nodes" with very few miR partners each. Examination of the degree distribution in the miR combinatorial network revealed a power law with a slope of about −1.5 and $R^2 = \sim 0.89$ in TargetScan and $R^2 = 0.94$ in PicTar (Figures 3B and S4B), indicating that the network of coregulating miRs is scale-free (alternative FDR cutoffs also resulted in scale-free networks with R^2 always bigger than 0.72). Interestingly, expression data of the miRs provides some support for the predicted regulatory interactions between them. We found that coexpressed miRs tended to have relatively high co-occurrence scores, and significant co-occurrence *p*-values, while miR pairs with negatively correlated expression tended to avoid residing in shared 3′ UTRs (see below).

Coordinated Regulation of Target Genes by miRs and TFs

A potential regulatory design in the gene expression network is that genes belonging to the same regulon will be coregulated not only at the transcriptional level, but also posttranscriptionally [31]. One potential realization of this design may be that a particular miR and a particular TF would regulate common targets. A simple means to identify some of the cases of regulatory cooperation between a miR and a TF may be to find TF–miR pairs that co-occur in a large set of shared targets compared with the size expected by chance. Similar to the case of miRs sites in 3′ UTRs, we considered a TF to be present in a human gene's promoter only if its occurrence in the promoter is conserved in the promoters of orthologous genes from mouse and rat [32] (as taken from UCSC, see Materials and Methods). We then created a matrix whose rows are the genes and columns are TFs, with a "1" for the i-th gene and the j-th TF if the TF

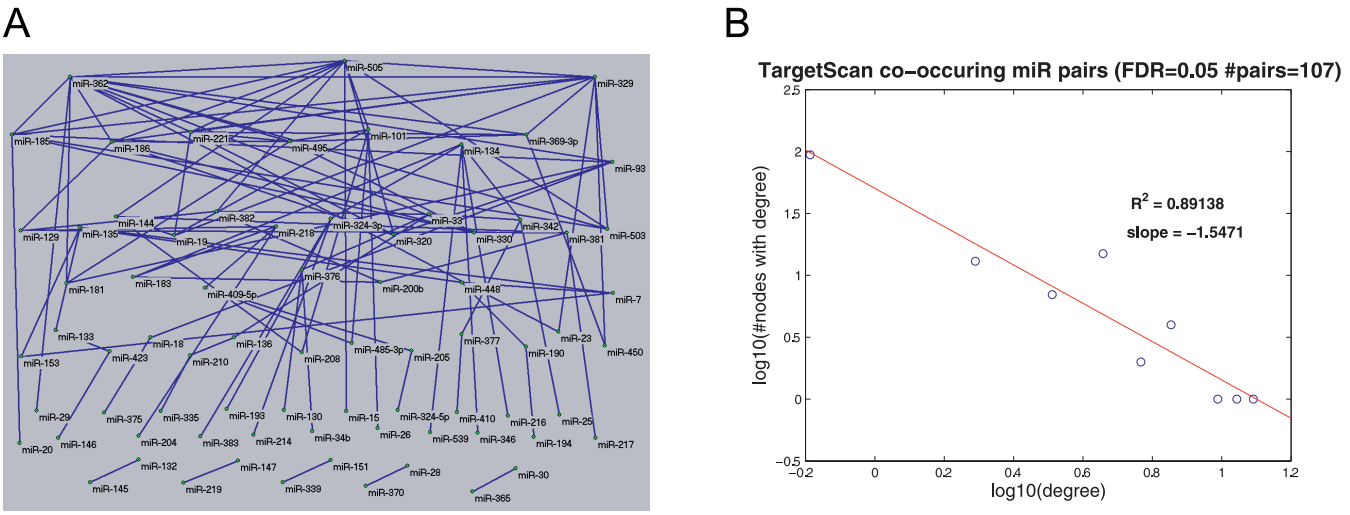


Figure 3. miR Co-Occurrence Network in the TargetScan Dataset

(A) The TargetScan miR co-occurrence network, at FDR level of 0.05. A node represents a miR and an edge connects between pairs of miRs with significant rate of co-occurrence. The nodes in the figure are arranged from most highly connected on the top, to most lowly connected, on the bottom. For interactive viewing of the network, using Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/), see Datasets S1 and S2.

(B) Degree distribution in the TargetScan miR combinatorial regulation network (co-occurring miR pairs that passed FDR of 0.05).

doi:10.1371/journal.pcbi.0030131.g003

binding site (TFBS) occurs in the gene's promoter and "0" otherwise.

To identify pairs of TFs and miRs that cooperate in regulating shared target genes, we looked for TF–miR pairs with a high rate of co-occurrence in the promoters and 3′ UTRs of the regulated genes. We tested the co-occurrence in shared genes of each of the 409 position specific scoring matrices (PSSMs) representing TF binding sites in TRANSFAC [13] with each of the 138 and 178 miRs in the TargetScan and PicTar databases, respectively. A PSSM and a miR are said to co-occur in the same gene if the PSSM has a conserved binding site in the promoter of the gene and the miR has a conserved predicted site in the gene's 3′ UTR. We used two statistical models to calculate the significance of rate of TF–miR co-occurrence, and ultimately considered TF–miR pairs that were found to be significant according to both tests. First, a hypergeometric p-value was calculated based on the number of genes that contain a TFBS in their promoter, the number of genes that contain a miR site in their 3′ UTR, and the number of genes that contain both the TF and the miR sites (see Materials and Methods for details). We computed such p-values on all TF–miRs pairs and set a threshold on the p-values obtained to account for the multiplicity of hypotheses, using FDR. Using an FDR q-value of 0.3, we obtained 111 miR-TF pairs with significant p-values using the TargetScan dataset and 1,263 miR-TF pairs with significant p-values using the PicTar dataset (see Materials and Methods for number of pairs with more stringent q-values). Reassuringly, there is a high overlap between the TargetScan and PicTar networks (68.7% of the TargetScan miR–TF network pairs were also found to be significant pairs in the PicTar network). The hypergeometric p-value has the advantage of being an analytical model with essentially unlimited resolution. Also, unlike the above situation of miR co-occurring pairs, which exhibited inherent dependency between the two regulators, the present case of TF–miR interaction does not present such limitation (and is in fact identical to the classical cases in which hypergeometric model is used [33]). Nevertheless, we decided to also back up the hypergeometric-based predictions with a randomization test, very similar to the one presented above for the case of miR co-occurrence, that preserves the distribution of number of regulators of each gene, the number of targets of each TF, and the number of targets of each miR in the input datasets. We calculated the co-occurrence rates and p-values of all TF–miR pairs, and used FDR as above to account for the multiplicity of hypotheses (see Materials and Methods for details). Reassuringly, 93% and 72% of the hypergeometric-based TF–miR interactions from the TargetScan and PicTar datasets, respectively, were also supported by this alternative model. The rest of the analyses were based on TF–miR pairs that passed the two statistical tests using FDR; there were 104 pairs in the TargetScan dataset and 916 pairs in the PicTar dataset. For simplicity we term a TF and a miR that significantly co-occur as partners. Table 3 lists the top TF–miR partners. The full networks of TF–miR partners can be downloaded as Tables S3–S5, and interactively viewed in Datasets S3–S5.

The Network of miR–TF Coregulation Reveals Recurring Local Architectures—Network Motifs

Recently it was suggested that in circuits composed of a miR and a TF, in which these two regulators target the same genes, the TF may also exert a regulatory effect on the miR with which it coregulates the target genes [22]. It was suggested that such a feed-forward loop (FFL) [19,20], a well-known local feature of many biological networks, may have a beneficial function. An FLL consisting of a TF and a miR could act as a switch for developmental and other programs in cells, since it may acquire biological systems with robustness to noise by means of canalization of perturbations [22]. We wanted to check whether in any of the significant miR-TF partners discovered above, the miR and its partner TF may regulate each other. We determined how many of the miR–TF partner pairs (out of 104 pairs in the TargetScan

Table 3. Top 20 Most Significant Pairs of Coregulating miRNAs and TFs in the TargetScan and PicTar Networks

Dataset	miR	Number of Target Genes	TF	Number of Target Genes	Intersection Size	Hypergeometric Co-Occurrence *p*-Value	z-Score
TargetScan	miR-7	209	V$NRSF_01	141	21	6.80E-08	7.96
	miR-153	301	V$NRSF_01	141	24	5.94E-07	4.80
	miR-381	444	V$CHX10_01	176	34	1.38E-06	4.49
	miR-133	254	V$NRSF_01	141	21	1.84E-06	5.64
	miR-448	302	V$NRSF_01	141	23	2.35E-06	4.77
	miR-369–3p	295	V$CDPCR1_01	87	17	3.05E-06	4.35
	miR-221	176	V$CDPCR1_01	87	13	3.52E-06	5.60
	miR-135	292	V$NRSF_01	141	22	4.79E-06	5.63
	miR-323	327	V$NKX25_02	284	37	4.86E-06	3.89
	miR-103	271	V$E47_02	11	6	5.46E-06	9.44
	miR-362	88	V$TCF11MAFG_01	132	11	7.49E-06	5.91
	miR-505	191	V$CDPCR1_01	87	13	8.70E-06	5.20
	miR-362	88	V$CDPCR1_01	87	9	9.29E-06	6.09
	miR-142–5p	319	V$GATA1_05	141	22	1.99E-05	4.09
	miR-324–5p	70	V$PAX6_01	153	10	2.18E-05	6.30
	miR-362	88	V$P53_01	75	8	2.40E-05	7.76
	miR-324–5p	70	V$HOX13_01	47	6	2.61E-05	7.02
	miR-503	202	V$PAX6_Q2	92	13	2.96E-05	6.55
	miR-23	410	V$NKX25_02	284	41	3.09E-05	3.35
	miR-129	163	V$GFI1_01	132	14	3.48E-05	3.79
PicTar	hsa-miR-199a,b*	151	V$CDPCR1_01	89	15	2.23E-09	8.58
	hsa-miR-326	261	V$NRSF_01	154	25	3.30E-09	8.14
	hsa-miR-28	124	V$AR_02	45	10	1.15E-08	9.26
	hsa-miR-369–3p	328	V$CDPCR1_01	89	20	1.97E-08	6.81
	hsa-miR-153	359	V$NRSF_01	154	27	1.32E-07	5.95
	hsa-miR-139	236	V$POU1F1_Q6	177	23	1.56E-07	6.01
	hsa-miR-185	267	V$PPARG_01	93	17	2.85E-07	6.52
	hsa-miR-199b	190	V$TCF11MAFG_01	164	19	4.21E-07	5.60
	hsa-miR-328	207	V$ARP1_01	86	14	6.66E-07	6.48
	hsa-miR-7	252	V$NRSF_01	154	21	7.32E-07	5.86
	hsa-miR-186	415	V$CDPCR1_01	89	20	9.45E-07	5.08
	hsa-miR-320	382	V$CDPCR1_01	89	19	1.16E-06	5.90
	hsa-miR-142–5p	284	V$GATA1_05	166	23	1.34E-06	5.19
	hsa-miR-132	191	V$TEL2_Q6	85	13	1.42E-06	5.66
	hsa-miR-212	191	V$TEL2_Q6	85	13	1.42E-06	6.32
	hsa-miR-133a,b*	333	V$NRSF_01	154	24	1.57E-06	5.89
	hsa-miR-155	178	V$EVI1_04	156	17	1.68E-06	6.15
	hsa-miR-139	236	V$GATA2_01	18	7	2.21E-06	6.14
	hsa-miR-302a,d*	342	V$ISRE_01	177	26	2.99E-06	5.46
	hsa-miR-135a	325	V$NRSF_01	154	23	3.63E-06	5.37
	hsa-miR-10a	131	V$POU6F1_01	302	20	3.74E-06	5.44

List of the top 20 most significant pairs of miR–TF coregulators. The *p*-value is a hypergeometric *p*-value for the co-occurrence of a miR and a TF in the 3′ UTRs and promoters of the same genes, and the *z*-score is assigned according to the randomization based co-occurrence method. The table depicts the number of targets of each miR and each TF, and the number of targets which contain sites for both miR and TF.

*In the PicTar table, the pairs of duplicated miRs (a, b, c, etc.) were unified when they appeared more than once as significant. The details (number of genes and *p*-value) presented in these unified rows are the details for the most significant pair of the assembly.

doi:10.1371/journal.pcbi.0030131.t003

dataset and 916 pairs in the PicTar dataset) had a conserved TF binding site of the partner TF in the putative upstream regulatory region of the partner miR (see Materials and Methods for definition of miRs' upstream putative regulatory regions). Interestingly, we found that ten of the TF–miR pairs in the TargetScan dataset (9.6% of the pairs), and 75 out of 916 pairs in the PicTar dataset (8.2%) fulfilled that additional requirement (see Figure 4). To establish whether this rate was significant, we carried out a randomization test (see Materials and Methods) in which we computed, in 10,000 randomized sets of TF–miR pairs, the rate of formation of a regulatory interaction between the TF and the miR. In the TargetScan network, we obtained a modest *p*-value of 0.024; however, in both PicTar networks we obtained the minimal possible *p*-value, $<10^{-4}$, i.e., in all 10,000 randomizations we got a rate of direct regulatory interaction between a TF and the miR, which was lower than the original data (see corresponding z-scores in Figure 4). Thus, the cases in which a TF and a miR co-occur in a highly significant number of target genes was associated more often than random with a direct regulation between the TF and the miR's promoter. We named this feed-forward loop "FFL TF → miR." The significance of this motif is robust to "noise" in the input, assessed by the method originally used for network motifs in *Escherichia coli* [20] (see Materials and Methods).

We were also interested in the opposite interaction—i.e., the case in which the miR regulates its partner TF. We named this motif "FFL miR → TF." We determined how many of the miR–TF partners had a predicted binding site of the partner miR in the 3′ UTR of the partner TF; it occurred five times in

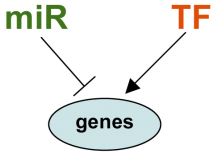
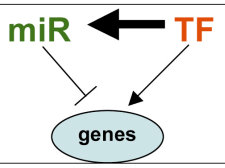
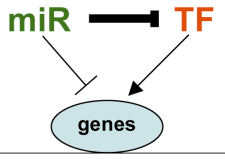
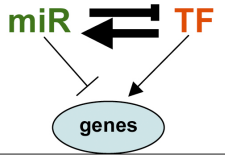
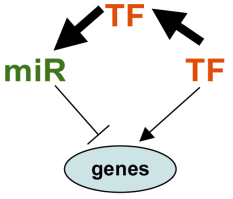
Motif		Number of appearances		p-value	Z-score	Number of genes	example
Significant pair		TargetScan	104	$6*10^{-8}$-$5*10^{-4}$* ($<10^{-3}$)	2.49-10.15**	1017	miR-1 & V$CREBP1CJUN_01
		PicTar 5kb	497	$1*10^{-7}$-$2*10^{-3}$* ($<10^{-3}$)	2.53-21.42**	1959	
		PicTar 10kb	916	$2*10^{-9}$-$5*10^{-3}$* ($<6*10^{-3}$)	2.13-13.49**	3336	
Feed Forward Loop (FFL) TF➔miR		TargetScan	10	0.0255	2.36	182	V$IRF2_01 ➔ miR-10
		PicTar 5kb	45	$<10^{-4}$	5.3	493	
		PicTar 10kb	75	$<10^{-4}$	5.58	861	
FFL miR➔TF		TargetScan	5	0.14	1.39	94	miR-148 ➔ V$FOX_Q2
		PicTar 5kb	42	$<10^{-4}$	6.03	275	
		PicTar 10kb	48	0.0038	2.93	455	
Composite loop miR➔➔TF		TargetScan	0	-	-	-	Cluster miR-25, miR-93, miR-106b & V$E2F: E2F ➔miR-93 & All three ➔ E2F
		PicTar 5kb	6	$4*10^{-4}$	5.27	76	
		PicTar 10kb	7	0.0024	4.02	116	
Indirect FFL		TargetScan	30	0.0013	3.32	485	V$SRF_Q6 ➔V$KROX_Q6 and V$NGFIC_01 ➔ miR-19b
		PicTar 5kb	95	$<10^{-4}$	4.55	855	
		PicTar 10kb	201	$<10^{-4}$	4.28	1635	

Figure 4. Network Designs in the miR–TF Coregulation Network

The figure depicts the analyzed network motifs in the TargetScan and PicTar dataset, and with the use of TF binding sites in RefSeq genes promoters of 10 kb for both networks, and 5 kb for the PicTar network. The figure depicts, for each network motif, its architecture, the number of times it appears in each of the networks, the *p*-value and z-score for its over-representation in the network (as described in Materials and Methods), the total number of RefSeq genes that are regulated by this type of network design, and an example.

*For the first design, the coregulating miR–TF pair, we state the range of hypergeometric *p*-values of pairs that passed FDR and are considered significant, and in brackets the FDR *p*-value of these pairs using the randomization co-occurrence test.

**In addition, z-scores for significant pairs were calculated based on the co-occurrence edge-swapping randomization model (see Materials and Methods).

doi:10.1371/journal.pcbi.0030131.g004

the TargetScan network, and 42 and 48 times in the PicTar networks, using two cutoffs on gene regulatory region lengths. This rate was not significant in the TargetScan network (p-value = 0.16), yet it was significant in the PicTar networks (p-values 0.0038 and $<10^{-4}$). Interestingly, we also found a composite loop network motif, which we termed "FFL miR ← → TF," in which the pair of partners regulate each other, to be significantly over-represented in the PicTar network; it appeared seven times in the PicTar network (see Figure 4).

In the next step, we looked for another type of network motif, that we termed an "indirect FFL," in which the TF's regulation on its partner miR is exerted via another mediator TF. We looked to see if any of the miR–TF partners in the network had a conserved TF binding site in a promoter of at least one other TF, which in turn has a conserved binding site in the promoter of the partner miR. Significantly, this architecture was very common in our networks; 30 of the TF–miR partners in the TargetScan network (28%) and 201 partners in the PicTar network (22%) were connected in a regulatory path between the TF and the miR via another TF. We tested the significance of these results by a randomization test, similar to that described above (see Materials and Methods), and received a p-value of 1.3×10^{-3} for the appearance of the indirect FFL in the TargetScan network, and p-value $< 10^{-4}$ for the PicTar network (see Figure 4). For the full list of motifs see Tables S3–S8.

Expression Analyses Supports miR–TF and miR–miR Predicted Regulatory Interactions

We next analyzed the expression profiles of TF–miR partners. Expression data across human tissues and organs has recently become available for miRs [34] and is also available for protein coding mRNAs [35]. Fortunately, for all the five healthy tissues (brain, liver, thymus, testes, and placenta) for which miRs expression was assayed, mRNAs were measured too. We could thus calculate the correlation coefficient between the expression profiles of each mRNA

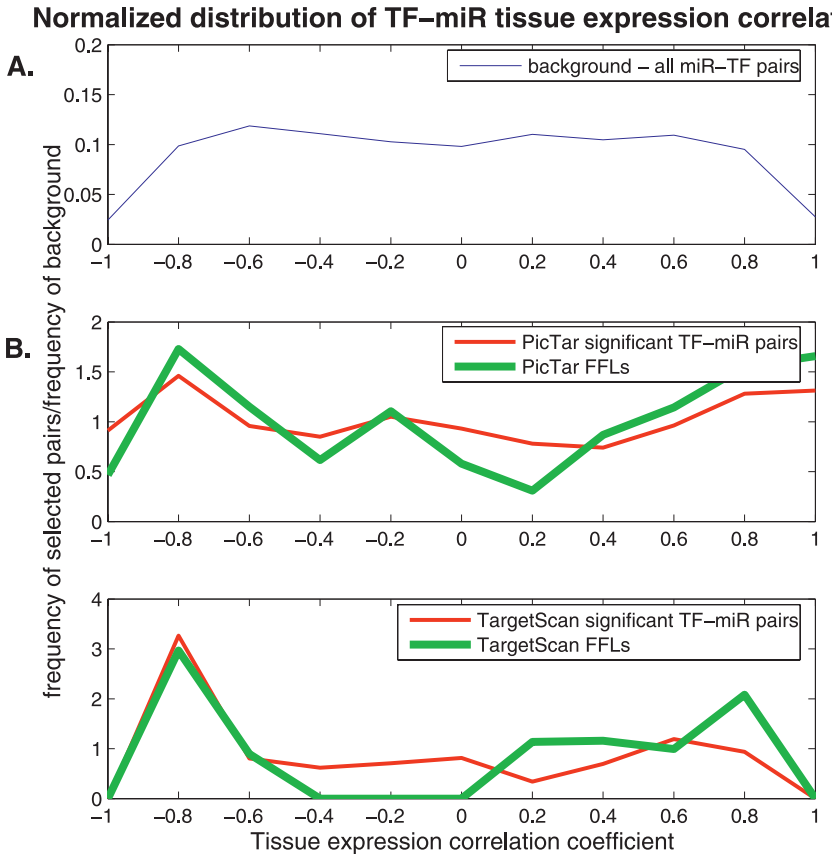


Figure 5. Tissue Expression Correlations between miRs and TFs

miR tissue expression in brain, liver, thymus, testes, and placenta were taken from [34]. mRNA tissue expression was taken from [35].
(A) Background distribution of all possible miR–TF pairs for which expression profiles can be derived.
(B,C) Normalized histograms of correlation coefficients; the same distribution as in (A) was made, yet only for significantly co-occurring miR-TF pairs (red), and FFLs (green) in the PicTar (B) and TargetScan (C) networks. The figure shows the proportion of the various correlation coefficients divided by the background distribution depicted in (A).
doi:10.1371/journal.pcbi.0030131.g005

and each miR, and in particular between all TF–miR partners. For background statistics, we first calculated correlations between all pairs of miRs and TFs in the expression dataset (i.e., not necessarily the TF–miR partners identified above) and obtained their distribution, and found, as may be expected, a distribution that is centered on zero (Figure 5A). On this background we show the distribution of correlation coefficients between expression profiles of TF–miR partner pairs (Figure 5B and 5C). Strikingly, we found that TF–miR partner pairs tended to have high correlation coefficients between them, but, curiously, there was also a tendency for strong negative correlations in some of these pairs. These two tendencies were further enhanced when we inspected only the TF–miR pairs that are connected through an FFL. Given that some TFs can act as activators and others as repressors, and given that miRs may act at the level of translation inhibition or transcript degradation, both negative and positive correlations between TF–miR partners may be mechanistically rationalized.

We further used the same miR tissue expression data to shed light on the co-occurrence and avoidance of miR pairs. We tested whether pairs of miRs that are either highly correlated in their expression levels or anticorrelated to each other across human samples have particularly high co-occurrence or avoidance p-values. We found an encouraging correspondence, whereby miR pairs that were positively correlated in expression had a significant tendency for high co-occurrence, whereas miRs with negative correlation in tissue expression typically tended to deliberately avoid residing in shared 3′ UTRs (Figure S5). These observations provide experimental support for miR pairs and TF–miR regulatory interactions that were initially predicted based on sequence information alone.

Discussion

We provide here a comprehensive characterization of both global and local structural properties of the network of combinatorial regulatory interactions spanned by miRs and TFs. We discovered extensive interactions between miRs and between miRs and TFs, and realize that thousands of human genes are subject to their regulatory effects. Inspection of the distributions of predicted miR sites in human genes' 3′ UTRs revealed hundreds of target hubs [24] in the human genome, genes that appear to be controlled by multiple regulators—miRs in the present case. Curiously, the current target hubs show highly nonrandom representation of specific gene functionalities. Particularly, genes related to development and genes that regulate transcription are enriched among the set of target hubs. These findings constitute another

demonstration of the recent concept [24] that suggests that genes that exert extensive regulation on crucial processes are themselves often heavily regulated. So far this has been discussed in the context of the yeast transcription network; this study extends the scope of this concept to the case of miRs in mammalian genomes. In addition, given that each method of target prediction has its own rate of false positives, target hubs, which are predicted to be targeted by multiple miRs, are more likely to actually represent true targets of miR silencing.

The network of extensive regulatory interactions observed here between transcriptional regulators (TFs) and post-transcriptional regulators (miRs), is another interesting global feature. Altogether we estimate that the number of human genes that are under combined regulation at the transcriptional and posttranscriptional silencing levels is between ~1,000 and ~4,000 (i.e., ~12% to ~43% of the ~9,000 analyzed genes, according to the TargetScan and PicTar networks, respectively). Overall, ~9,000 genes were included in the present analyses. These are genes that are currently predicted to have at least one binding site for a known miR. Considering the fact that the collection of mammalian miRs is yet incomplete, and the fact that human specific miRs were not included in the analysis, we anticipate that the true number of human genes that are subject to a dual TF–miR regulation were underestimated in this study. For comparison, we recently estimated that in the *Saccharomyces cerevisiae* genome about 13% of the genes are subject to regulation at the combined transcriptional and posttranscriptional level [31], albeit with different mechanisms of posttranscriptional regulation operating in this organism, which does not have the miR silencing pathway.

We also examined local properties of the regulatory network, the network motifs. The network motifs described here are different from those originally described [18–20] in that they are composed of a TF and a miR instead of two TFs, as in the original case. We have shown here that network motifs are not only significantly abundant, but also that, according to their current definition, each of them is involved in the regulation of a large set of targets. Interestingly, TF and miR pairs that participate in network motifs show a significant tendency toward high tissue expression correlations or anticorrelations of the two regulators, providing essential experimental support to combinations predicted solely based on sequence information.

Motifs in which the miR regulates its partner TF constitute a type II coherent FFL [18]. In this case it seems that a miR that silences a set of genes posttranscriptionally also silences the transcriptional regulator of these genes, presumably to also prevent de novo transcription of its target genes. This design may be used to minimize leaky transcription of genes in space and time when their expression is undesired. For example, this mechanism could be useful in determining developmental fate in differentiation boundaries as also suggested by [22,23,36].

The motifs in which the TF has a binding site in the promoter of its partner miR corresponds to the incoherent type I FFL (assuming that the TF is a positive regulator). Interestingly, in the *S. cerevisiae* transcription network this circuit is the second most highly abundant FFL [18]. An intriguing question is what may be the reason for the observed abundance of this circuit in which a TF regulates its partner miR? On the face of it, such regulation appears wasteful if the TF is a positive regulator, since the TF activates an entire set of genes and also a miR that may shut down those target genes. However, if a temporal gap in the activation time of the target genes and the miR exists, then the circuit may be utilized for useful regulatory purposes. For instance, if the TF activates first the target genes and only later the miR (e.g., due to higher affinity, [20]), during a process in which the TF's concentration builds up, the activation of the miR may be timed to obtain a desired delayed shutdown of the regulated genes. We have recently considered similar wiring in the cases of antisense RNAs, another type of regulatory transcripts, and TFs that regulate them in conjunction with their overlapping sense transcripts [37]. The opposite situation, in which the TF positively activates the miR first and only later the target gene, may also be of interest as it can act as a buffer for noisy fluctuations in the levels of the targets; as long as the mRNA level of the target gene is below the inhibition capacity of the miR, fluctuations in its expression levels would not be further propagated. Further, in cases where the miR works predominantly as a translation inhibitor, a controlled mechanism for "just in time" translation for multiple genes is needed for certain functionalities. For example, the miR translation inhibition mechanism was suggested to facilitate localized translation in mammalian dendrites, and to play a crucial role in synaptic plasticity [38]. Such a circuit of coregulating TF–miR in an FFL, where the miR is transcribed by the TF in parallel to the set of mutual targets, could function in featuring localized translation to a whole pathway of regulated genes. Interestingly though, we can point out an example of one indirect FFL we discovered, where a brain-related TF, CREB (CREBATF) [39], partners with a miR that is known to be expressed in the brain, miR-125b [40]. CREBATF was predicted by us to regulate miR-125b through STAT3, which interestingly is also within the list of mutual targets of both miR125b and CREBATF, indicating an even more complex design.

One of the FFLs that came out of our analysis is a composite loop in which the TF regulates the miR and the miR appears to regulate the TF (i.e., a TF $\leftarrow \rightarrow$ miR motif). The circuit consists of the TF E2F and miR-93. miR-93 is part of a cluster of three miRs, miR-106b, miR-93, and miR-25, which lie in close proximity to each other inside an intron of the *MCM7* gene. This network motif was found as an FFL TF $\rightarrow$ miR in the TargetScan network and as a composite loop in the PicTar network, where all three miRs in the cluster were predicted to target E2F (specifically E2F1 and E2F3). miR-93 cluster members are also homologous to two other genomic miR clusters, one of which is miR cluster 17/92 [41]. Recent evidence suggests a tight regulatory connection of cluster miR-17/92 and E2F [42–45]. E2F1, 2, and 3 were shown to directly upregulate the expression of the miRs encoded in this cluster, while these miRs in turn were shown to act in a feedback loop and to target E2F1–3 mRNAs [42,43]. It was suggested that this feedback may play a role in the major decision mediated by E2F (induction of cellular proliferation or apoptosis). Here we would like to suggest that this intricate regulatory circuit might have another layer to it; in addition to being targeted by the miR-17/92 cluster, E2F family genes might also be targeted by miR-93 cluster members, which share similar seeds. In turn, the miR-93 cluster is transcribed from an intron of the *MCM7* host gene, which is a verified target of the E2F family [46]. Moreover, here the architecture

is more complex, as it also includes a set of mutual target genes, through which E2F and the miR-93 cluster may exert their regulatory roles.

Future experimental work will allow the examination of the predictions generated here and the establishment of their precise regulatory roles.

Materials and Methods

miRs and their predicted targets. miRs and their predicted targets were taken from two previously published studies: TargetScan [8,9] (http://www.targetscan.org) and PicTar [7] (http://genome.ucsc.edu). Both resources predict and assign target genes to miRs based on evolutionary conservation between human, mouse, rat and dog. TargetScan targets were downloaded 21 September 2006 and gene symbols were converted to RefSeq IDs using UCSC mysql databases. PicTar targets were downloaded 25 September 2006 from the UCSC hg17 database [7,32] where they are presented as the picTarMiRNA4Way track.

Target hubs analysis. Target hubs were defined as genes which are targeted by more miRs than the 99th percentile of the maximal value in 100 randomizations of the columns in the miR to gene assignment matrix; each preserved the total number of targets per miR. According to this procedure, in the TargetScan dataset, target hubs were defined as genes which are targeted by more than 15 miRs (there were 470 such genes), and in the PicTar dataset, target hubs were defined as genes targeted by more than 20 miRs (834 genes). For original and randomized distributions see Figures 1A and S1A.

We wanted to check whether the target hubs contain many miR target sites merely because they have, on average, longer 3′ UTRs. For that purpose, the length of 3′ UTRs for all RefSeq genes was retrieved from UCSC hg17. We performed a randomization test on this 3′ UTR length data, in which we randomly picked a set of genes from the data with distribution of 3′ UTR length that was as similar as possible (see below) to that of the target hubs. For each such set of genes we calculated the average number of different miRs predicted to target them. We repeated this randomization procedure 100 times, and the distribution of average number of miRs was derived (Figure S3). The figure shows that these values are significantly lower than the average of the real target hubs, indicating that the length is neither necessary nor sufficient for a gene to be a target hub.

We generated 100 random sets of genes with length distributions similar to that of the target hubs by the following procedure. For each target hub with UTR length, L_{TH}, we defined a set of genes with similar UTR length, which included all the genes in the dataset with a UTR length equal to L_{TH}, or longer up to an additional 5% of L_{TH} (genes which did not have such sets were excluded from the analysis). Then, we randomly chose a representative from each set to be included in the randomized version of target hubs. miR density in the 3′ UTRs of genes was calculated as the number of miRs targeting a gene divided by its 3′ UTR length. The 3′ UTR length was extracted from the UCSC database.

When defining high density target hubs we chose the density cutoffs to be the top 85th percentile of the entire distribution of densities. We note that this distribution included only genes that participated in our analyses and thus does not contain genes with a density of zero (i.e., zero predicted sites in the UTR).

Degree-preserving matrix randomization. To determine a p-value on the co-occurrence rate of a pair of two miRs, we first defined a co-occurrence score. We chose the Meet/Min score [29,30], which is formulated in the main text, and calculated it on the matrix of miR to target genes. For the purpose of p-value calculations we defined a null model of randomized matrices, which preserves the matrix statistics such that for each gene the number of miRs targeting it, and for each miR the number of genes it targets remains the same as in the original data. This model was first introduced as a randomization model for networks [20], which preserved all in and out degrees in a given network, and thereby controlling for the possibility that significance of a phenomenon may be merely attributed to the degree distribution in the network. Randomized matrices were created by the edge-swapping procedure, starting from the original matrix of miR to target gene predictions. We randomly picked two pairs of miR and target gene, miR_{i1}–$gene_{j1}$ and miR_{i2}–$gene_{j2}$, and, after verifying that miR_{i1} does not already target $gene_{j2}$ and miR_{i2} does not already target $gene_{j1}$, we performed the switch of an edge in the matrix, so that after the swap there is a "0" instead of "1" in the positions i_1,j_1 and i_2,j_2 in the matrix, and a "1" instead of a "0" in the positions i_1,j_2 and i_2,j_1 in the matrix. To decide how many swapping events were needed before the matrix was "well randomized," we monitored the number of edges that were actually swapped and compared it with the number of changed edges in a randomly shuffled matrix. We followed this number during the swapping steps and realized that it plateaued at about 100,000 steps. Thus, in all subsequent analyses we repeated the swapping procedure for 100,000 steps.

During the calculation of the Meet/Min score for a pair of miRs in the original data, we excluded genes that contained a match to the two miRs if the two sites physically overlapped on the target's 3′ UTR. In addition, we filtered out from the analysis pairs of miRs whose seeds were identical (overlap of seven out of seven nucleotides, positions 2–8 of the miR). These two precautions were taken to eliminate the possibility of overestimating the significance of the rate of miR co-occurrence due to seed sequence similarity between different miRs.

After having calculated the co-occurrence p-values and avoidance p-values for all possible miR pairs, we controlled for multiple hypotheses using FDR and only pairs that passed FDR of 0.05 were considered to be significantly co-occurring or avoiding.

Significant miR–TF co-occurring pairs. For the task of identifying miR–TF pairs that significantly co-occur in a high number of target genes, a p-value was calculated (using a cumulative hypergeometric test) on each pair of regulators as we did before for pairs of TFs [14]. The hypergeometric p-value was calculated after the RefSeq genes were mapped to a unique set of Gene IDs, to reduce redundancy in the set. In the miR–TF p-value calculations, the total number of genes in the hypergeometric analysis was calculated as the number of genes that appeared (i.e., had at least one binding site) in both datasets. Genes that appeared only in the TF dataset or in the miR dataset were excluded and were not counted. We used FDR to correct for multiple hypotheses testing, and determined the set of significant pairs of coregulators.

We also calculated co-occurrence p-values for all possible miR–TF pairs using the new randomization method presented above. Specifically, both the matrix which assigns TFs to genes and the matrix with assignments of miRs to genes were subjected to 100,000 iterations of the edge-swapping procedure. In total we generated 1,000 such pairs of randomized matrices. The co-occurrence p-value of a given TF–miR pair is the fraction of the randomized matrix pairs in which this pair's Meet/Min score was higher than the pair's Meet/Min score in the original matrices, and the corresponding z-score is the difference between the original Meet/Min score and the mean of the score in the randomized matrices, divided by their standard deviation.

Most reassuringly, when checking the overlap of these significant pairs with the significant pairs that passed FDR cutoff of 0.3 using the hypergeometric model, we saw that the overlap was very high; it was more than 72% for PicTar and 92% for TargetScan. For subsequent analyses of network motifs (FFLs and indirect FFL search), we chose all the pairs that passed FDR of 0.3 in the hypergeometric test in the three datasets (see Transcription factor binding sites section below), and that passed FDR of 0.3 (p-value $< 6 \times 10^{-3}$) in the PicTar 10 kb set, and minimal p-value ($< 10^{-3}$) in the PicTar 5 kb and TargetScan sets, as these already had an extremely high overlap (>93%) in the hypergeometric derived set.

The final set of significant pairs in the miR–TF network is presented in FDR q-value cutoffs of 0.1, 0.2, and 0.3. With q-value of 0.1 we obtained 20 TF–miR pairs with significant p-value using the TargetScan dataset, and 267 using the PicTar 10 kb dataset, and 70 using the PicTar 5 kb dataset. With a q-value of 0.2 we obtained 60 TF–miR pairs with significant p-value using the TargetScan dataset, and 555 using the PicTar 10 kb dataset, and 261 using the PicTar 5 kb dataset. With 0.3 we obtained 104 TF–miR pairs with significant p-value using the TargetScan dataset, and 916 using the PicTar 10 kb dataset, and 497 using the PicTar 5 kb dataset.

miRs clusters and regulatory regions. As was shown in the past [41], miRs may be clustered on the genome, and are often transcribed as one unit. Therefore, to predict regulatory regions of miRs (i.e., proximal as well as potentially more distant promoters or enhancers) we had to first cluster miRs on the human genome. We mapped all 461 pre-miRs in miRBase (http://microrna.sanger.ac.uk, accessed June 2006) [47,48] onto the human genome and clustered them according to physical proximity (genomic locations of miRs were taken from UCSC hg17 and some miRs were mapped from hg18 back to hg17 using the UCSC "lift genome" web service). Two pre-miRs, that are consecutive on the genome, were considered belonging to the same cluster if the distance between them was shorter than a cutoff, provided that they are transcribed from the same strand. We kept adding miRs to clusters until we hit the first distance that was larger than the cutoff. To learn a meaningful cutoff from the data, we plotted the distribution of distances between all neighboring pre-

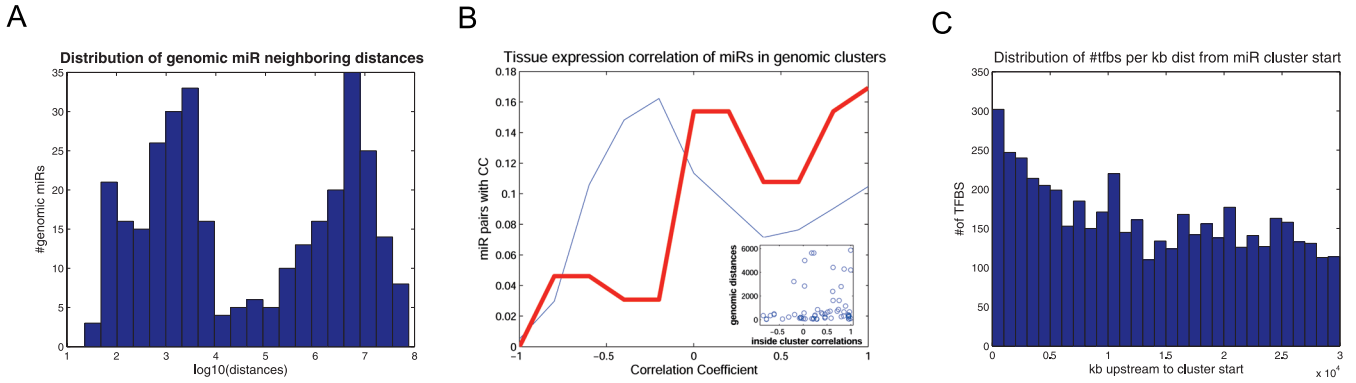


Figure 6. Analysis of miR Clusters in the Human Genome

(A) Distribution of distances between all neighboring pre-miR genes in the human genome.

(B) Distribution of tissue expression correlations between pairs of miRs: all possible pairs in the data (thin blue line) and pairs of miRs which reside in shared clusters (thick red line). In the inset are shown tissue expression correlations between pairs of miRs in the same genomic clusters versus distances between them.

(C) Distribution of number of conserved TFBS 30 kb upstream of the 5′ most nucleotide in each miR clusters. Conserved TFBSs were taken from UCSC hg17.

doi:10.1371/journal.pcbi.0030131.g006

miRs in the genome. Interestingly, we found the distribution to be bimodal—distances below and above 10 kb (on a log scale, Figure 6A) were highly represented in contrast to a lower representation at about 10 kb. This indicated that a reasonable cutoff on the distance between two adjacent miRs that still belong to the same cluster may be 10 kb. Using this clustering procedure we generated 301 clusters, the majority of which (~82.39%) consists of a single miR; the cluster with the highest number of miRs contains 43 miRs (see Figure S7 for the distribution of number of miRs per cluster). In a previous study, which was based on 207 miRs (compared with the 461 used here), miRs were clustered using a different cutoff [49]. When we repeated our cluster analysis with the current set of miRs, with the previous cutoff, we got similar clustering, 94% of the present clusters are identical to the clusters generated with the alternative cutoff and average cluster lengths are very similar (unpublished data).

Reassuringly, using expression data of miRs across tissues [34] we found that miRs that belong to the same cluster have a significant tendency to be coexpressed compared with miRs that do not map to shared clusters (Figure 6B). This tendency is preserved even in cases where miRs that belong to the same cluster are relatively far from each other on the genome (Figure 6B, inset).

We have then defined, as a putative regulatory region of miRs, the sequence that lies 10 kb upstream of the 5′ most pre-miR in each miR cluster. The 10 kb promoter length was determined from the data as follows. A distribution of number of conserved TFBS upstream of clusters was generated (Figure 6C). We found that the number of conserved TFBS gradually declined as a function of the distance from the putative 5′ end of the cluster, with a plateau obtained at about 10 kb upstream. The distribution was rather noisy, probably due to the fact that primary-miR transcripts are much longer than the precursor miR we relate to (e.g., the primary transcript of the miR-17–92 cluster is C13orf25, which is 6,795 bp long [45]), and thus the transcription start site (TSS) taken here is only crudely defined. We considered the presence of a TFBS in a miR promoter only if such occurrence was conserved in mouse and rat, as taken from the UCSC hg17 conserved track in the relevant regions.

Transcription factor binding sites. We used predicted binding sites for all human mouse and rat PSSMs from TRANSFAC [13] version 8.3, as they are defined by the UCSC hg17 genome assembly, in the tfbsConsSites (http://genome.ucsc.edu/) and tfbsConsFactors. All RefSeq genes genomic locations were taken from hg17. To determine the length of upstream regulatory regions, we measured the number of conserved TFBS upstream RefSeq genes as a function of distance from TSS (see Figure S6). The result shows that the signal decays and plateaus between 5 kb and 10 kb upstream of the TSS. We hence chose to work with two alternative cutoffs of promoter length, 5 kb and 10 kb. The regulatory regions thus defined probably consist of proximal promoters as well as distant enhancers. The recent Affymetrix (http://www.affymetrix.com) promoter chip for detection of ChIP experiments with TF binding in human promoters also consists of probes that span 10 kb of regulatory regions, and future experiments with this chip and as many TFs as possible will allow a better delineation of regulatory regions boundaries. Although we used regulatory regions which are longer than the common definition, our use of evolutionary conservation filter gives confidence in the present regulatory region definitions.

Feed-forward loop statistics. FFL TF → miR: for all the significant pairs of coregulators (i.e., TF–miR partners that co-occur in a significantly high number of targets) we investigated whether the TF has a binding site in the putative promoter of the miR cluster from which the miR partner is transcribed. In some cases in which the mature miR sequence is transcribed from more than one genomic locus, all possible regulatory regions of the relevant miR clusters were examined. In addition, each PSSM may belong to a family of PSSMs, with similar binding sites, representing the same TF (a family was defined as several PSSMs representing the same TF, as determined from the UCSC hg17 tfbsConsFactors track). Thus, PSSM–miR pairs are treated as TF–miR pair, and given a pair of PSSM–miR partners, we say that the PSSM's TF regulates the miR if at least one of the PSSMs that corresponds to that TF has a match in the regulatory region of the miR partner (the same procedure was carried out in the randomizations described below).

For testing the FFL miR → TF configuration, we had to connect first between TRANSFAC PSSMs and the genes encoding the TFs that bind these PSSMs. For that, PSSMs were mapped to the TF they represent which in turn was mapped to a SwissProt ID. These two mappings were done using the UCSC hg17 tfbsConsFactors track. These SwissProt IDs were then mapped to RefSeq IDs, for which the data on miR targets was maintained. This information served also in the process of indirect FFL search; for each of the TF–miR partners, we checked whether the miR is regulated by another mediator TF, which in turn is regulated by the partner TF. We note that not all TFs had a corresponding SwissProt ID in the UCSC hg17 tfbsConsFactors track, and therefore not all pairs served as candidates for the FFL miR → TF and the indirect FFL; only in 74 of the 104 (71%) TargetScan significant pairs, and in 680 of 916 (74%) of the PicTar pairs, could the PSSM be mapped to a RefSeq gene.

The following procedure was used for the calculation of the significance of the FFLs and indirect FFL in the PicTar and TargetScan miR–TF networks. Since there were 104 and 916 pairs of miR–TF partners in the two respective networks, we have drawn 10,000 times the same number of random pairs of TFs and miRs out of all the possible pairs in each network. The number of each FFL and indirect FFL was recorded in each randomization and a p-value (and a corresponding z-score) on the hypothesis that a given network motif is over-represented in the network was taken to be the number of random sets with a greater or equal number of motifs in it.

miR and mRNA tissue expression data. The expression profiles of 150 miRs across five healthy human tissues and organs (brain, liver, thymus, testes, and placenta) were previously measured using miR-dedicated microarrays [34]. miRs from the chips were mapped to PicTar and TargetScan; they cover 154 and 87 of the miRs in the two respective datasets. In addition, we used data from [35] for human mRNAs expression across the same set of tissues. Both sets of expression data were column centered (chip-wise centering: each chip's values were divided by the chip mean to account for differences

in chip intensities) and then log2 transformed. Regarding mRNA expression chips, we particularly focused on genes coding for the TFs that participated in our analysis. Using the above mapping of PSSMs to their corresponding TF genes, we had a total of 127 TFs that could be matched to at least one probe set in the mRNA expression dataset [35]. We examined the tissue expression correlation of all significantly co-occurring miR and TF pairs for which we had an expression profile. When more than one gene was attributed to the same TF, we chose for each pair of TF and miR the one with the highest absolute value of correlation coefficient out of all options. We did that consistently both for the background statistics of all possible TF–miR pairs and for the predicted TF–miR partners. In total we calculated correlation coefficients for 361 such TF–miR partners out of 916 partners in PicTar, and for 30 out of 104 partners in TargetScan. The miR expression data [34] consisted of five healthy tissues, and HeLa cells, while the mRNA study that we focused on [35] overlapped with the miR data only in the five tissues. Therefore when we compared expression between miRs and TFs we only used the five healthy tissues, and when we compared expression of miR pairs we used all six samples.

Noise-tolerance analysis. The assignments of miRs to targets are known to be of limited accuracy [21] . We thus wanted to assess the noise tolerance of our results. We adopted a procedure previously utilized for the case of network motifs in the bacterial transcription network [20]. We experimented with different percentages of the connections in the network that were randomly removed or added and the significance of the present FFL motifs was assessed for each case. Similarly to the findings in the *E. coli* network, we found that up to 20%–30% of the edges can be added or removed without appreciable effect on the FFL significance.

Supporting Information

Dataset S1. Pajek Input File for the miR Co-Occurrence Network, the TargetScan Dataset (Significant Co-Occurring miR Pairs with FDR *q*-Value 0.05)

All networks in the Dataset files can be interactively viewed using the Pajek software, which can be freely downloaded from (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

Found at doi:10.1371/journal.pcbi.0030131.sd001 (12 KB TXT).

Dataset S2. Pajek Input File for the miR Co-Occurrence Network, the PicTar Dataset (Significant Co-Occurring miR Pairs in FDR *q*-Value 0.05)

Found at doi:10.1371/journal.pcbi.0030131.sd002 (20 KB TXT).

Dataset S3. Pajek Input File for the Network of miR–TF Coregulating Pairs

This graph depicts all the significant miR–TF pairs in the TargetScan network, in addition to all the FFLs. A red node is a TF and a green node is a miR, and a blue edge is drawn if the TF and the miR are co-occurring partners. A yellow edge connects between a TF and a miR if, in addition to having a high rate of co-occurrence, they also form a FFL TF → miR; a pink edge represents the FFL miR → TF motif, while orange edge represents a FFL miR ← → TF (in all cases the set of target genes is not explicitly shown).

Found at doi:10.1371/journal.pcbi.0030131.sd003 (16 KB TXT).

Dataset S4. Pajek Input File for the Network of miR–TF Coregulating Pairs

This graph depicts the 100 most significant pairs in the PicTar (10 kb) network, in addition to all the FFLs.

Found at doi:10.1371/journal.pcbi.0030131.sd004 (86 KB TXT).

Dataset S5. Pajek Input File for the Network of miR–TF Coregulating Pairs

This graph depicts the 100 most significant pairs in the PicTar (5 kb) network, in addition to all the FFLs.

Found at doi:10.1371/journal.pcbi.0030131.sd005 (55 KB TXT).

Figure S1. Distribution of miRs to Target Gene Assignments in the PicTar Dataset

(A) Distribution of the number of different miRs regulating each target gene in the PicTar dataset. The thick red line represents the distribution in the original datasets, while each of the thin blue lines represents the distribution in one of the column-randomized matrices. The matrix contains only genes with at least one predicted site in their 3′ UTR. In each randomization, we shuffle the assignment of miRs to their targets, keeping constant the number of targets per miR.
(B) Distribution of number of targets per miR in the PicTar dataset. In the thick red line we depicted the original distribution, while each blue thin line represents the distribution in one of the 100 row-randomized matrices, which preserve the distribution of number of miRs targeting each gene.

Found at doi:10.1371/journal.pcbi.0030131.sg001 (1.6 MB EPS).

Figure S2. miR Binding Sites and 3′ UTR Length in the TargetScan and PicTar Datasets

A dot plot depicting number of miRs targeting each gene and its 3′ UTR length of the target hubs, high miR number target hubs in green, high density target hubs in red, genes that are target hubs according to both criteria in magenta and the rest of the genes in blue for the (A) TargetScan dataset and (B) PicTar Dataset.
(C) Distribution of the miR densities in the 3′ UTRs of target hubs (thick red line) and all the genes (thin blue line) in the PicTar dataset (all genes included in this figures have at least one miR site predicted in their 3′ UTR). The log10 densities were binned into bins of 0.1, and relative frequencies were plotted.

Found at doi:10.1371/journal.pcbi.0030131.sg002 (1.6 MB EPS).

Figure S3. miR Binding Sites in Target Hub Genes in the TargetScan and PicTar Datasets
Mean number of miRs targeting each of the genes that are target hubs (red bar), in the entire set of analyzed genes (green), and a distribution of that mean in random gene sets with the same (or very similar, see Materials and Methods) distribution of 3′ UTR lengths as the target hubs (blue) in (A) the TargetScan dataset and (B) the PicTar dataset. For elaborated procedure see Materials and Methods.

Found at doi:10.1371/journal.pcbi.0030131.sg003 (1.6 MB EPS).

Figure S4. miR Pairs Interaction Network in the PicTar Dataset

(A) The miR pairs interaction network in the PicTar database.
(B) Degree distribution in the PicTar miR combinatorial regulation network (co-occurring miR pairs that passed FDR of 0.05)

Found at doi:10.1371/journal.pcbi.0030131.sg004 (1.6 MB EPS).

Figure S5. Positively Correlated miR Pairs Tend To Have Significant Co-Occurrence *p*-Values while Negatively Correlated Pairs Tend to Avoid Residing in the Same 3′ UTRs

Highly expression correlated miR pairs tend to have significant co-occurrence or *p*-values, while negatively correlated pairs tend to have significant avoidance *p*-values. The figures depict the Kolmogorov-Smirnov *p*-values for the hypotheses that correlated miR pairs have lower co-occurrence *p*-values than the rest of the pairs. Correlated pairs were defined according to correlation cutoffs (depicted on the *x*-axis), with positively correlated pairs in blue, negatively correlated pairs in green. Positively correlated miR pairs tend to have significant co-occurrence *p*-values in both TargetScan (A) and PicTar (C). Negatively correlated pairs tend to have significant avoidance *p*-values in both TargetScan (B) and PicTar (D).

Found at doi:10.1371/journal.pcbi.0030131.sg005 (3.9 MB EPS).

Figure S6. Distribution of Number of Conserved TFBS 30 kb Upstream of TSS of RefSeq Protein-Coding Genes

Found at doi:10.1371/journal.pcbi.0030131.sg006 (11 KB EPS).

Figure S7. Distribution of Number of miRs per Cluster

As seen, ~82% of the 301 clusters contain a single miR.

Found at doi:10.1371/journal.pcbi.0030131.sg007 (12 KB EPS).

Table S1. Significant Co-Occurring miR Pairs in the TargetScan Dataset

Found at doi:10.1371/journal.pcbi.0030131.st001 (30 KB XLS).

Table S2. Significant Co-Occurring miR Pairs in the PicTar Dataset

Found at doi:10.1371/journal.pcbi.0030131.st002 (38 KB XLS).

Table S3. Significant Co-Occurring miR–TF Pairs in the TargetScan Network

Found at doi:10.1371/journal.pcbi.0030131.st003 (32 KB XLS).

Table S4. Significant Co-Occurring miR–TF Pairs in the PicTar Network, Taking 10 kb Regulatory Regions for Protein Coding Genes

Found at doi:10.1371/journal.pcbi.0030131.st004 (172 KB XLS).

Table S5. Significant Co-Occurring miR–TF Pairs in the PicTar Network, Taking 5 kb Regulatory Regions for Protein Coding Genes

Found at doi:10.1371/journal.pcbi.0030131.st005 (103 KB XLS).

Table S6. Indirect FFLs in the TargetScan Dataset

Found at doi:10.1371/journal.pcbi.0030131.st006 (22 KB XLS).

Table S7. Indirect FFLs in the PicTar Dataset Taking 10 kb Regulatory Regions for Protein Coding Genes

Found at doi:10.1371/journal.pcbi.0030131.st007 (47 KB XLS).

Table S8. Indirect FFLs in the PicTar Dataset Taking 5 kb Regulatory Regions for Protein Coding Genes

Found at doi:10.1371/journal.pcbi.0030131.st008 (29 KB XLS).

Acknowledgments

We thank E. Hornstein, D. Ginsberg, and M. Lapidot for stimulating discussions and for critical review of the manuscript. Y. P. is an incumbent of the Rothstein Career Development Chair in Genetic Diseases. R. S. is a Fellow of the Horowitz Foundation for Complexity Sciences.

Author contributions. YP and RS conceived and designed the project. RS wrote the computer code and performed the experiments; RS, DL, MO, and YP analyzed data. YP and RS wrote the paper.

Funding. We thank the Ben May Charitable Trust for grant support.

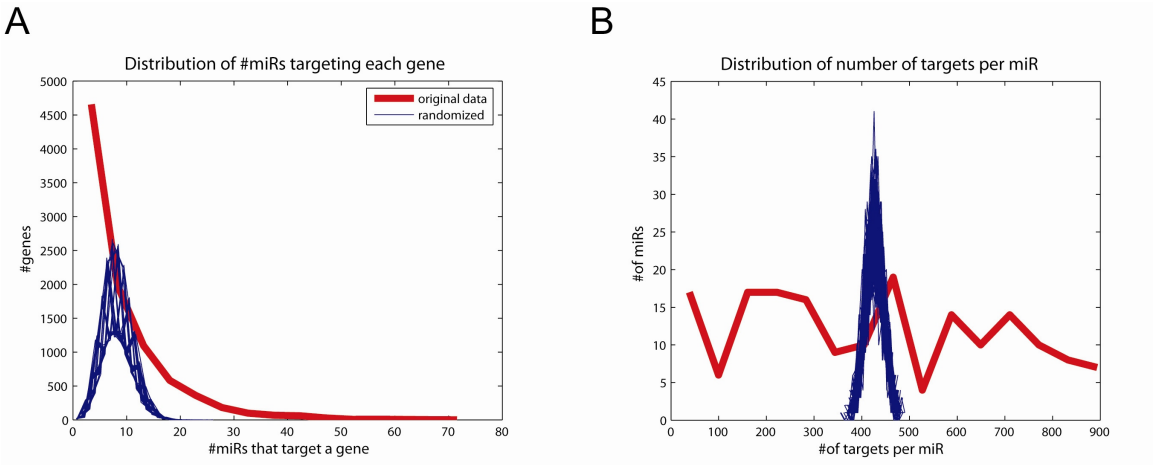
Competing interests. The authors have declared that no competing interests exist.

References

1. Bartel DP (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell 116: 281–297.
2. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37: 766–70.
3. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T (2003) New microRNAs from mouse and human. RNA 9: 175–179.
4. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, et al. (2002) Identification of tissue-specific microRNAs from mouse. Curr Biol 12: 735–739.
5. Enright AJ, John B, Gaul U, Tuschl T, Sander C, et al. (2003) MicroRNA targets in *Drosophila*. Genome Biol 5: R1.
6. John B, Enright AJ, Aravin A, Tuschl T, Sander C, et al. (2004) Human MicroRNA targets. PLoS Biol 2: e363.
7. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. Nat Genet 37: 495–500.
8. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15–20.
9. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115: 787–798.
10. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, et al. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433: 769–773.
11. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434: 338–345.
12. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431: 99–104.
13. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, et al. (2003) TRANSFAC: Transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31: 374–378.
14. Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 29: 153–159.
15. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, et al. (2003) Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet 34: 166–176.
16. Yu H, Gerstein M (2006) Genomic analysis of the hierarchical structure of regulatory networks. Proc Natl Acad Sci U S A 103: 14724–14731.
17. Kalir S, Mangan S, Alon U (2005) A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. Mol Syst Biol 1: 2005.0006.
18. Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 100: 11980–11985.
19. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: Simple building blocks of complex networks. Science 298: 824–827.
20. Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 31: 64–68.
21. Sethupathy P, Megraw M, Hatzigeorgiou AG (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. Nat Methods 3: 881–886.
22. Hornstein E, Shomron N (2006) Canalization of development by microRNAs. Nat Genet 38 (Supplement): S20–S24.
23. Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. Cell 123: 1133–1146.
24. Borneman AR, Leigh-Bell JA, Yu H, Bertone P, Gerstein M, et al. (2006) Target hub proteins serve as master regulators of development in yeast. Genes Dev 20: 435–448.
25. Yuh CH, Bolouri H, Davidson EH (1998) Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. Science 279: 1896–1902.
26. Garten Y, Kaplan S, Pilpel Y (2005) Extraction of transcription regulatory signals from genome-wide DNA-protein interaction data. Nucleic Acids Res 33: 605–615.
27. Sudarsanam P, Pilpel Y, Church GM (2002) Genome-wide co-occurrence of promoter elements reveals a *cis*-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. Genome Res 12: 1723–1731.
28. Elkon R, Linhart C, Sharan R, Shamir R, Shiloh Y (2003) Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. Genome Res 13: 773–780.
29. Goldberg DS, Roth FP (2003) Assessing experimentally derived interactions in a small world. Proc Natl Acad Sci U S A 100: 4372–4376.
30. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551–1555.
31. Shalgi R, Lapidot M, Shamir R, Pilpel Y (2005) A catalog of stability-associated sequence elements in 3′ UTRs of yeast mRNAs. Genome Biol 6: R86.
32. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, et al. (2003) The UCSC Genome Browser Database. Nucleic Acids Res 31: 51–54.
33. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. Nat Genet 22: 281–285.
34. Barad O, Meiri E, Avniel A, Aharonov R, Barzilai A, et al. (2004) MicroRNA expression detected by oligonucleotide microarrays: System establishment and expression profiling in human tissues. Genome Res 14: 2486–2494.
35. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101: 6062–6067.
36. Hornstein E, Mansfield JH, Yekta S, Hu JK, Harfe BD, et al. (2005) The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. Nature 438: 671–674.
37. Lapidot M, Pilpel Y (2006) Genome-wide natural antisense transcription: Coupling its regulation to its different regulatory mechanisms. EMBO Rep 7: 1216–1222.
38. Kim J, Krichevsky A, Grad Y, Hayes GD, Kosik KS, et al. (2004) Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci U S A 101: 360–365.
39. Lonze BE, Ginty DD (2002) Function and regulation of CREB family transcription factors in the nervous system. Neuron 35: 605–623.
40. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, et al. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol 5: R13.
41. Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. J Mol Biol 339: 327–335.
42. Woods K, Thomson JM, Hammond SM (2007) Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. J Biol Chem 282: 2130–2134.
43. Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, et al. (2007) An E2F/miR-20a autoregulatory feedback loop. J Biol Chem 282: 2135–2143.
44. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005) c-Myc-regulated microRNAs modulate E2F1 expression. Nature 435: 839–843.
45. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, et al. (2005) A microRNA polycistron as a potential human oncogene. Nature 435: 828–833.
46. Leone G, DeGregori J, Yan Z, Jakoi L, Ishida S, et al. (1998) E2F3 activity is regulated during the cell cycle and is required for the induction of S phase. Genes Dev 12: 2120–2130.
47. Griffiths-Jones S (2004) The microRNA Registry. Nucleic Acids Res 32: D109–D111.
48. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: MicroRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: D140–D144.
49. Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, et al. (2005) Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 33: 2697–2706.

Supplementary Figures

Figure S1



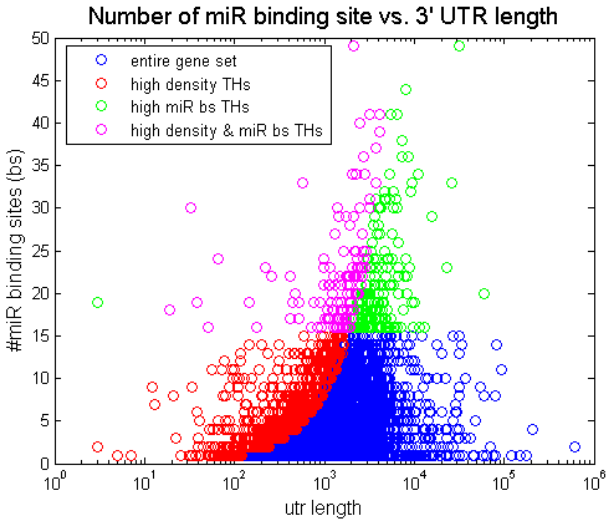
Distribution of miRs to Target Gene Assignments in the PicTar Dataset

(A) Distribution of the number of different miRs regulating each target gene in the PicTar dataset. The thick red line represents the distribution in the original datasets, while each of the thin blue lines represents the distribution in one of the column-randomized matrices. The matrix contains only genes with at least one predicted site in their 3' UTR. In each randomization, we shuffle the assignment of miRs to their targets, keeping constant the number of targets per miR.

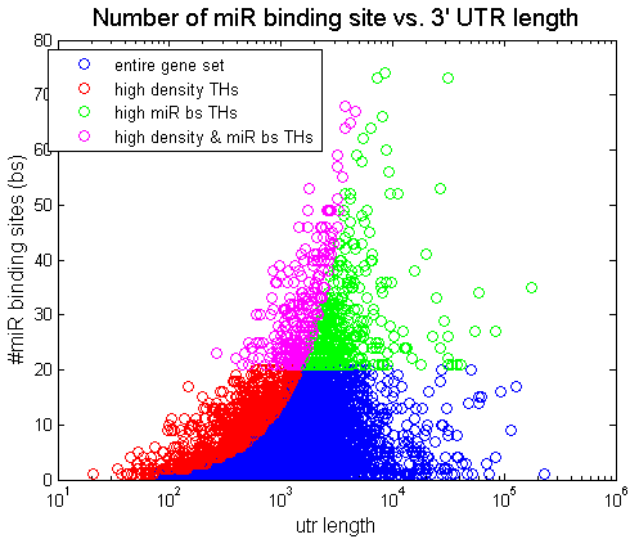
(B) Distribution of number of targets per miR in the PicTar dataset. In the thick red line we depicted the original distribution, while each blue thin line represents the distribution in one of the 100 row-randomized matrices, which preserve the distribution of number of miRs targeting each gene.

Figure S2

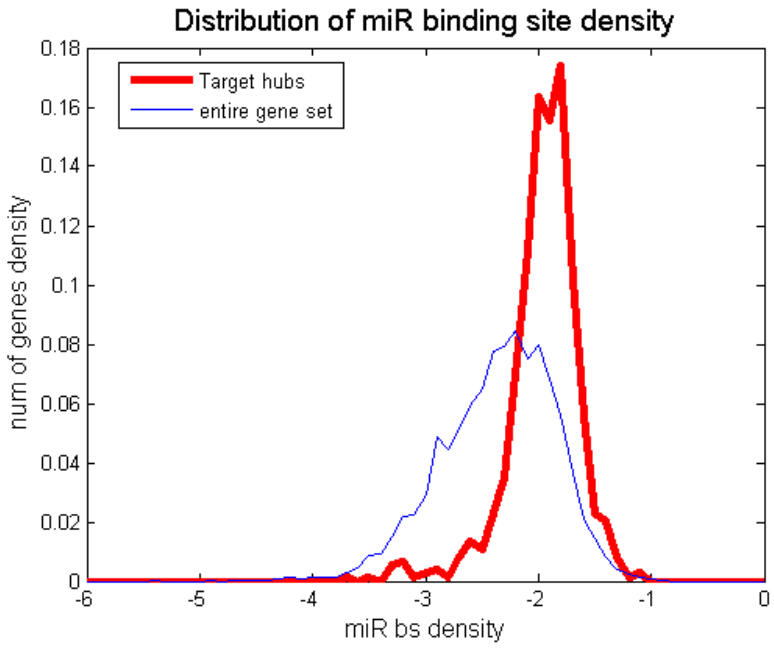
A.



B.



C.



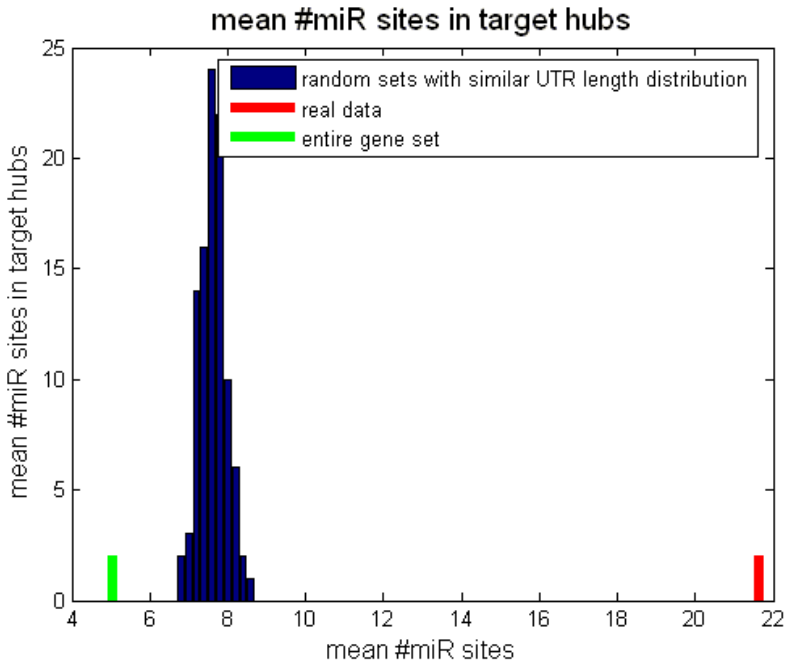
miR Binding Sites and 3′ UTR Length in the TargetScan and PicTar Datasets

A dot plot depicting number of miRs targeting each gene and its 3′ UTR length of the target hubs, high miR number target hubs in green, high density target hubs in red, genes that are target hubs according to both criteria in magenta and the rest of the genes in blue for the (A) TargetScan dataset and (B) PicTar Dataset.

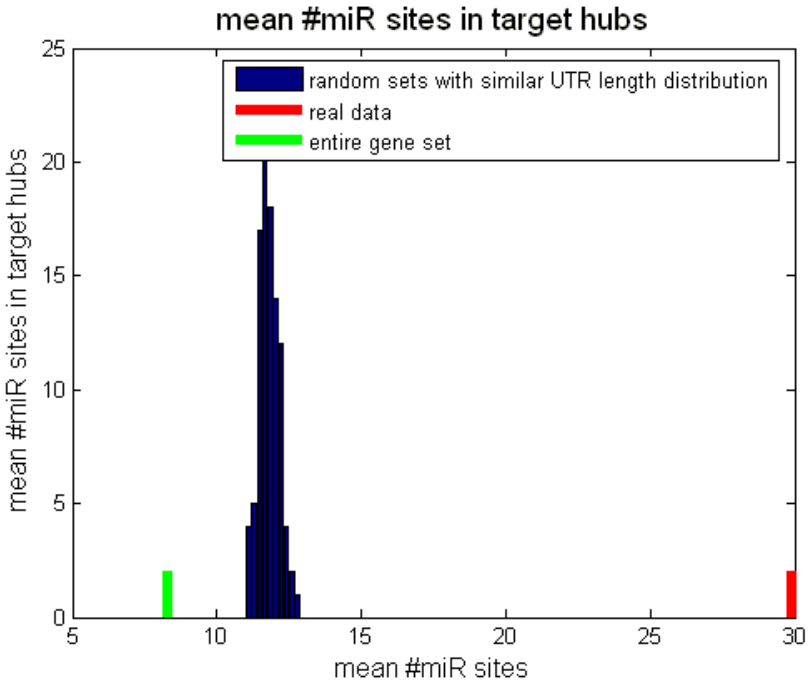
(C) Distribution of the miR densities in the 3′ UTRs of target hubs (thick red line) and all the genes (thin blue line) in the PicTar dataset (all genes included in this figures have at least one miR site predicted in their 3′ UTR). The log10 densities were binned into bins of 0.1, and relative frequencies were plotted.

Figure S3

A.



B.

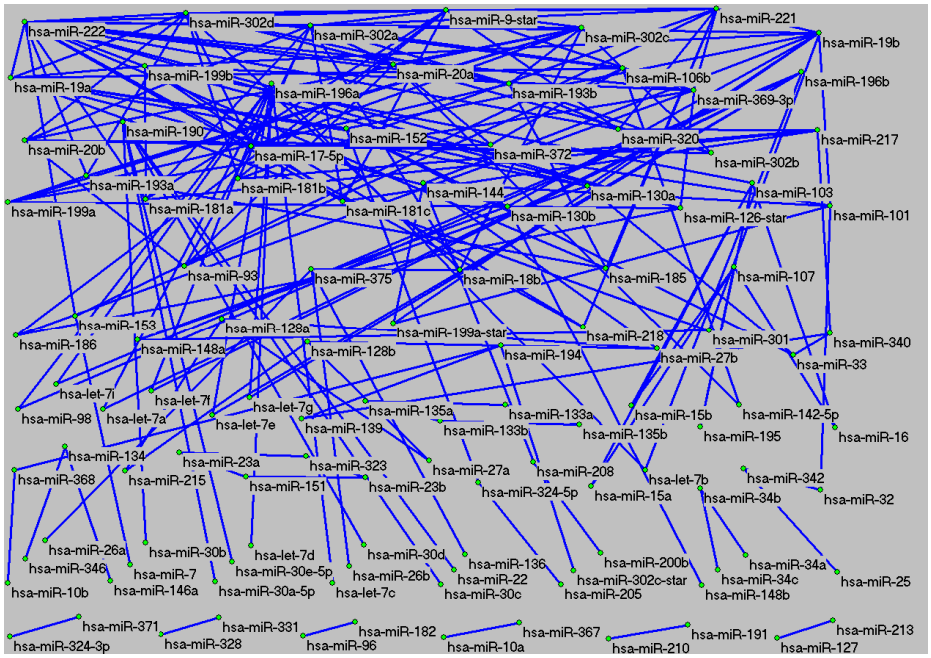


miR Binding Sites in Target Hub Genes in the TargetScan and PicTar Datasets

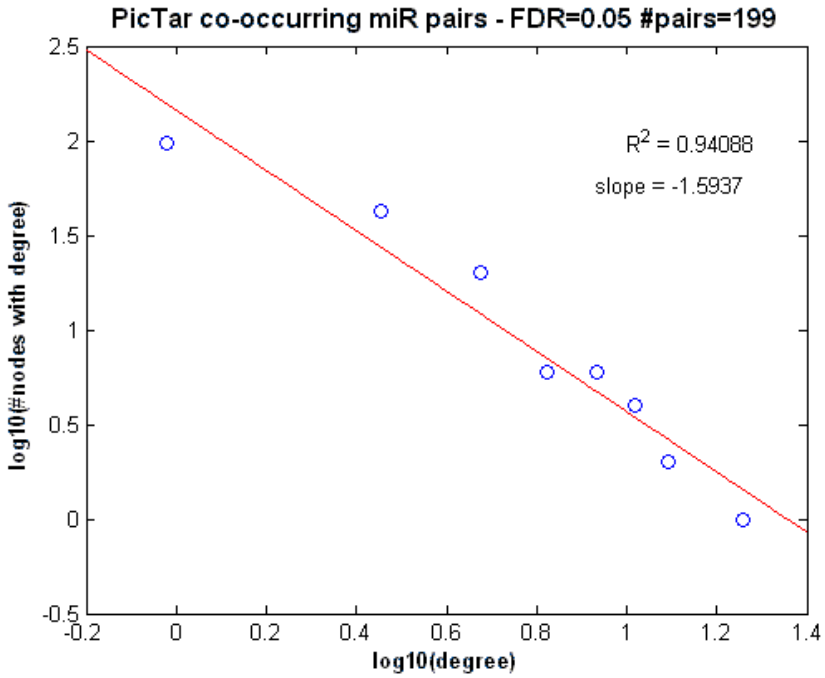
Mean number of miRs targeting each of the genes that are target hubs (red bar), in the entire set of analyzed genes (green), and a distribution of that mean in random gene sets with the same (or very similar, see Materials and Methods) distribution of 3′ UTR lengths as the target hubs (blue) in (A) the TargetScan dataset and (B) the PicTar dataset. For elaborated procedure see Materials and Methods.

Figure S4

A.



B.

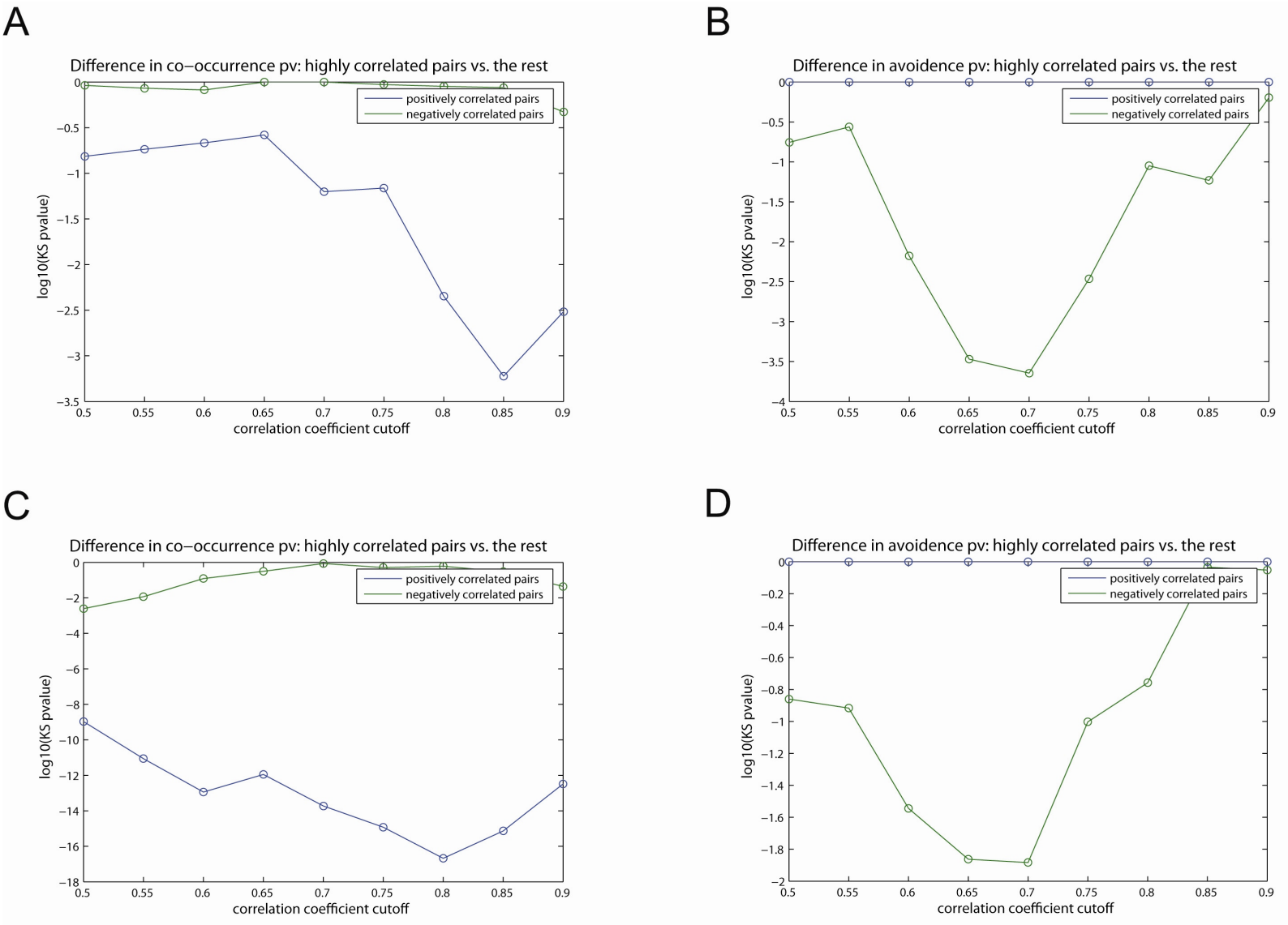


miR Pairs Interaction Network in the PicTar Dataset

(A) The miR pairs interaction network in the PicTar database.

(B) Degree distribution in the PicTar miR combinatorial regulation network (co-occurring miR pairs that passed FDR of 0.05)

Figure S5

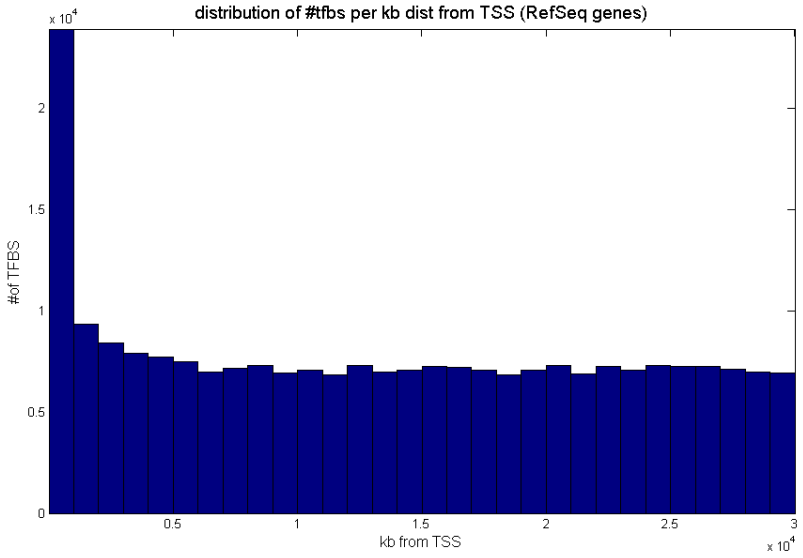


Positively Correlated miR Pairs Tend To Have Significant Co-Occurrence *p*-Values while Negatively Correlated Pairs Tend to Avoid Residing in the Same 3' UTRs

Highly expression correlated miR pairs tend to have significant co-occurrence or *p*-values, while negatively correlated pairs tend to have significant avoidance *p*-values. The figures depict the Kolmogorov-Smirnov *p*-values for the hypotheses that correlated miR pairs have lower co-occurrence *p*-values than the rest of

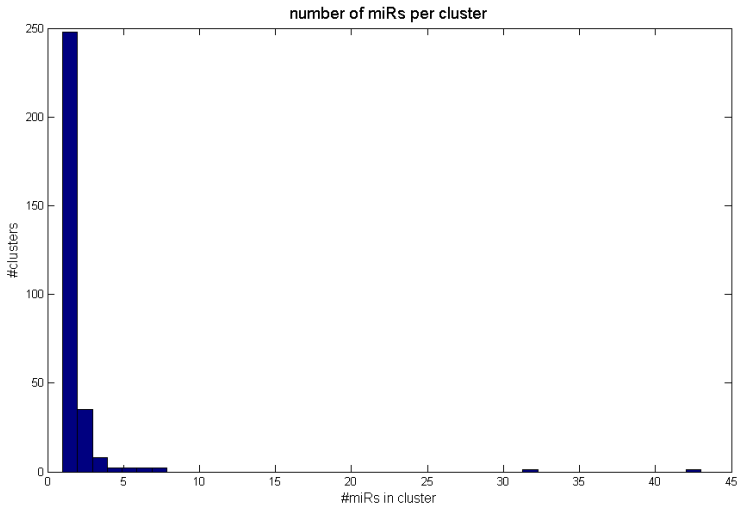
the pairs. Correlated pairs were defined according to correlation cutoffs (depicted on the *x*-axis), with positively correlated pairs in blue, negatively correlated pairs in green. Positively correlated miR pairs tend to have significant co-occurrence *p*-values in both TargetScan (A) and PicTar (C). Negatively correlated pairs tend to have significant avoidance *p*-values in both TargetScan (B) and PicTar (D).

Figure S6



Distribution of Number of Conserved TFBS 30 kb Upstream of TSS of RefSeq Protein-Coding Genes

Figure S7



Distribution of Number of miRs per Cluster

As seen, ~82% of the 301 clusters contain a single miR.

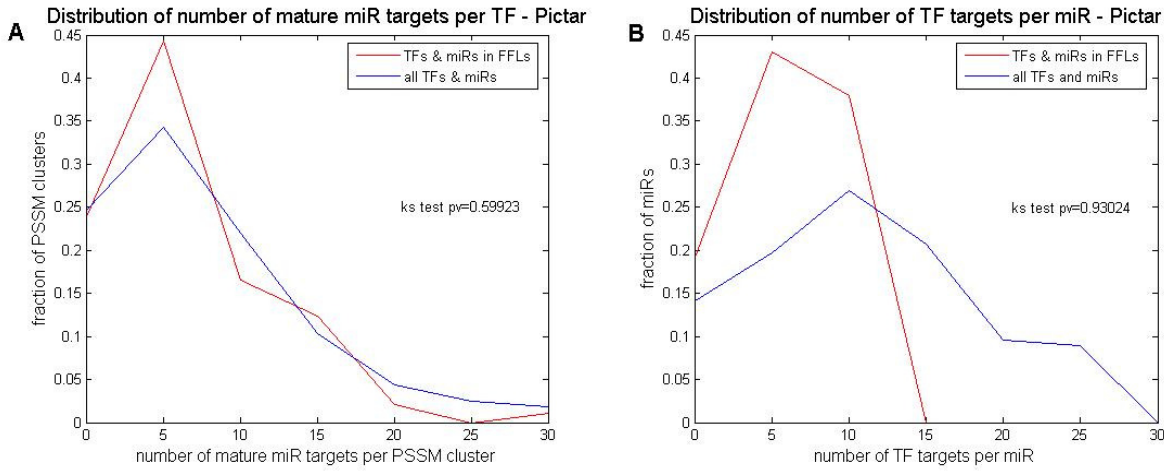
Additional control for degree distributions in FFLs

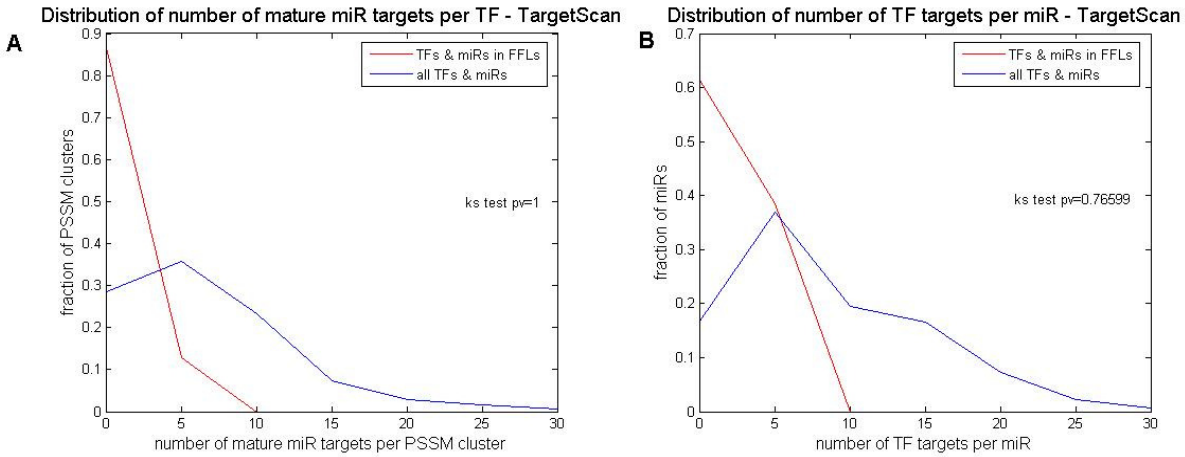
When looking for network motifs, i.e. for the over-representation of small architectures in a network, it is accustomed to control for the degree distribution of the nodes in the network. The degree distributions of miRNAs and TFs were taken into consideration in the edge-swap procedure, which was used to give a p-value on the significant pairs. In the procedure of giving an over-representation p-value on the third regulatory arrow which forms an FFL, i.e. miR→TF or TF→miR, we used a different statistical test, which utilizes random sampling. In order to ensure that the network motif p-values given on our different FFLs, are not a mere reflection of high degree TFs and miRs which happen to form significant pairs, we checked the degree distributions of those miRs and TFs, and compare them to the distribution of the entire set of regulators.

This is checked in four cases:

A. number of miR targets per TF (the same way which was used to determine the TF→miR FFLs)
B. number of TF targets per miR (the same way which was used to determine the miR→TF FFLs)

Here we demonstrate that in cases A. and B. the red distribution, which depicts the degree distributions only of the PSSMs and miRs participating in FFLs, is shifted to the left (towards lower degrees) compared to the general distributions (Kolmogorov-Smirnov tests on the hypothesis that the red distribution is larger than the blue show no difference, p-values indicated inside the figures). This is true for both PicTar and TargetScan. This fact ensures that there is no bias in our FFL network motif over-representation p-value, towards high degree regulators, and thus random sampling from the data, which is actually random sampling from a distribution of degrees which are higher than those in the original, is a satisfactory control for the general degree distribution in this network motif statistical test.





Comparison between the degree distributions of all miRs and PSSMs, and only those which were found to be a part of a Feed-Forward loop

in the figure we show the degree distribution of all miRs and PSSMs (blue lines), and the degree distribution of only those miRs and/or PSSMs which were found in the study to be a part of a Feed-Forward loop, either TF→miR of miR→ TF.

We show four different distributions: A. distribution of number of targets per miR in the Pictar dataset. B. distribution of number of target RefSeq genes per PSSM, taken from the UCSC dataset. C. distribution of number of mature miRs per TF – the dataset construction is explained in the Materials and Methods in the paper, under the section: **miRs clusters and regulatory regions**. D. distribution of the number of TF targets per miR in PicTar – the PSSMs were mapped to TFs, which in turn were mapped to RefSeq IDs and then, data about miR binding sites for those TFs was taken from PicTar, as described under the same section in the Materials and Methods.

For each such dataset, we derived the general distribution of number of targets per regulator, depicted in blue, and the distribution only for those miRs and TFs which were found to be part of FFLs TF→miR or miR→TF.

Molecular Systems Biology 4; Article number 229; doi:10.1038/msb.2008.65
Citation: ***Molecular Systems Biology*** **4:229**

www.molecularsystemsbiology.com

molecular systems biology

p53-repressed miRNAs are involved with E2F in a feed-forward loop promoting proliferation

Ran Brosh[1,8], Reut Shalgi[1,2,8], Atar Liran[1], Gilad Landan[1], Katya Korotayev[3], Giang Huong Nguyen[4,5], Espen Enerly[6,7], Hilde Johnsen[6], Yosef Buganim[1], Hilla Solomon[1], Ido Goldstein[1], Shalom Madar[1], Naomi Goldfinger[1], Anne-Lise Børresen-Dale[6,7], Doron Ginsberg[3], Curtis C Harris[4], Yitzhak Pilpel[2], Moshe Oren[1] and Varda Rotter[1,*]

[1] Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel, [2] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel, [3] The Mina and Everard Goodman, Faculty of Life Science, Bar Ilan University, Ramat Gan, Israel, [4] Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD, USA, [5] Howard Hughes Medical Institute-National Institute of Health Research Scholar, Howard Hughes Medical Institute, Chevy Chase, MD, USA, [6] Department of Genetics, Institute for Cancer Research, Norwegian Radiumhospital, Rikshospitalet University Hospital, Norway and [7] Department of Genetics, Faculty Division, The Norwegian Radiumhospital, University of Oslo, Oslo, Norway
* Corresponding author. Department of Molecular Cell Biology, Weizmann Institute of Science, Herzel, Rehovot 76100, Israel. Tel.: +972 893 440 70; Fax: +972 894 652 65; E-mail: varda.rotter@weizmann.ac.il
[8] These authors contributed equally to this work

Received 7.5.08; accepted 29.9.08

Normal cell growth is governed by a complicated biological system, featuring multiple levels of control, often deregulated in cancers. The role of microRNAs (miRNAs) in the control of gene expression is now increasingly appreciated, yet their involvement in controlling cell proliferation is still not well understood. Here we investigated the mammalian cell proliferation control network consisting of transcriptional regulators, E2F and p53, their targets and a family of 15 miRNAs. Indicative of their significance, expression of these miRNAs is downregulated in senescent cells and in breast cancers harboring wild-type p53. These miRNAs are repressed by p53 in an E2F1-mediated manner. Furthermore, we show that these miRNAs silence antiproliferative genes, which themselves are E2F1 targets. Thus, miRNAs and transcriptional regulators appear to cooperate in the framework of a multi-gene transcriptional and post-transcriptional feed-forward loop. Finally, we show that, similarly to p53 inactivation, overexpression of representative miRNAs promotes proliferation and delays senescence, manifesting the detrimental phenotypic consequence of perturbations in this circuit. Taken together, these findings position miRNAs as novel key players in the mammalian cellular proliferation network.
Molecular Systems Biology 25 November 2008; doi:10.1038/msb.2008.65
Subject Categories: bioinformatics; chromatin and transcription; molecular biology of disease
Keywords: breast cancer; microarray; miR-106; mir-155; senescence

Introduction

The tumor suppressor p53 is considered a central regulator of cell-fate decisions. Activation of p53 can induce several cellular responses, including cell-cycle arrest, senescence and apoptosis. Thus, absence of functional p53 predisposes cells to neoplastic transformation. Accordingly, mutations of this gene are highly common in human cancers (Hussain and Harris, 1999). p53 is a sequence-specific transcription factor (TF) that exerts many of its downstream effects by activating gene transcription (Ryan *et al*, 2001). Nevertheless, additional transactivation-independent functions of p53 contribute to its tumor suppressive activity, including protein–protein interactions with additional TFs and other cell-fate regulators. The importance of transcriptional regulation by p53 is exemplified by the fact that most p53 tumor-derived mutants are defective in DNA binding and incapable of transactivation (Kern *et al*, 1991). In addition to its capability to induce gene transcription, p53 activation results in extensive gene repression (Ginsberg *et al*, 1991). Direct and indirect transcriptional repression by p53 is considered important for its tumor suppressive functions, such as induction of cell-cycle arrest and apoptosis (Ho and Benchimol, 2003).

microRNAs (miRNAs or miRs in short) are a recently discovered class of small non-coding RNA species that regulate gene expression at the post-transcriptional level. Approximately half of the known miRs are encoded in regions of the genome that are distal to known genes, whereas the remaining

reside in introns, or in rare cases in exons, of coding genes, usually in the same orientation as the mRNA. Additionally, some miRs are clustered in the genome and are transcribed as polycistrons that may contain up to ~50 mature miRs (Bartel, 2004). With the recent identification of miRs that regulate cancer-related processes such as apoptosis, proliferation and differentiation, these RNA species emerge as important regulators of cancer initiation and progression. Accordingly, mutation and transcriptional deregulation of miRs have been linked to cancer (Esquela-Kerscher and Slack, 2006). Deregulated miRs were suggested to exert their function in cancer through silencing of key cell-fate regulators, as shown for let-7 and Ras (Johnson *et al*, 2005), as well as for miR-106b and p21 (Ivanovska *et al*, 2008; Petrocca *et al*, 2008).

In a previous work we suggested that miRs cooperate with certain TFs in the regulation of mutual sets of target genes, allowing the coordinated modulation of gene expression both transcriptionally and post-transcriptionally. Specifically, we found a recurring network motif in which a TF regulates the miR with which it cooperates in regulating a common set of targets, creating a feed-forward loop (FFL). One such case involved E2F and the miR-106b/93/25 polycistron (Shalgi *et al*, 2007).

Several studies have implicated p53 in the regulation of miR expression (Xi *et al*, 2006; Chang *et al*, 2007; He *et al*, 2007; Raver-Shapira *et al*, 2007; Tarasov *et al*, 2007; Kumamoto *et al*, 2008). These studies exploited various high-throughput methods to identify p53-regulated miRs in several cellular systems with differential p53 status. Although the resulting candidate lists from each study differed considerably, probably due to differences in the cellular contexts and p53 activation signals, all studies identified members of the miR-34 family as direct transactivation targets of p53. In line with p53 function, induction of miR-34 family members was suggested to mediate cell-cycle arrest, apoptosis and senescence. Importantly, none of these studies focused on miRs, whose expression negatively correlated with p53 activation, and which are likely repressed by this tumor suppressor.

Here, we report the identification of a large set of miRs, the expression of which constitutes a recurring signature in several experiments. The members of this signature are transcriptionally repressed by p53 in primary cells and in human breast cancers. This signature is comprised of known cancer-associated miRs as well as newly proposed ones, and includes the miR-106b/93/25 polycistron. We implicate E2F1 in the p53-dependent repression of these miRs, and demonstrate the oncogenic potential of the miR-106b/93/25 polycistron. Finally, we delineate a network architecture that includes the transcription factor E2F1 and a family of miRs, which co-regulate mutual target genes transcriptionally and post-transcriptionally, thereby enhancing cellular proliferation. This FFL is repressed by p53, possibly to promote senescence and suppress cancer progression.

Results

Identification of p53-regulated miRNAs in primary human cells and in human breast cancers

To identify novel p53-regulated miRs, we established two isogenic cell cultures that differ in their p53 status and analyzed their miRNA profiles both under normal conditions and in contexts involving p53 activation. WI-38 primary human fibroblasts were infected with a retrovirus encoding for the p53-inactivating peptide, GSE56 (Ossovskaya *et al*, 1996). These cells (GSE) and their active p53 counterparts (con) were treated with the DNA-damaging agent doxorubicin or grown until the onset of replicative senescence (the establishment of the system is depicted in Supplementary Figure S1). Analysis of miRNA expression patterns revealed several expression clusters (see Materials and methods). Notable among these was a cluster populated with miRs, the expression of which was negatively regulated by p53 under normal conditions. The cluster showed additional downregulation in senescent cells, which was attenuated upon p53 inactivation. We named this cluster the '*p53-repressed miR cluster*' (Figure 1A). Notably, doxorubicin treatment, which resulted in a considerable activation of p53 and its mRNA targets (Supplementary Figure S1), did not significantly affect the levels of these miRs.

Interestingly, a significant number of miRs from this cluster were also clustered together in a similar experiment, in which miR expression was profiled in young and senescent human embryonic fibroblasts (WI-38 and MRC5). This cluster is termed '*senescence-repressed miR cluster*' (Figure 1B). The significant overlap between the clusters (P-value=3.2×10^{-4}) is interesting as the second experiment was not designed to discover p53-regulated miRs, but rather to identify a general signature of miRs that are altered upon replicative senescence. However, p53 activity was increased in both senescent fibroblast cultures (data not shown). Strikingly, clustering analysis of miR expression data derived from a set of breast cancer tumors with differing p53 status also resulted in a cluster highly overlapping the '*p53-repressed miR cluster*' (Figure 1C; P-value=1.11×10^{-5}) (samples, p53 status and histological grading were described by Naume *et al* (2007) and Sorlie *et al* (2006), and detailed description of the mutation status is listed in Supplementary Table S1). The '*breast cancer p53-repressed miR cluster*' was comprised of miRs, the expression of which was negatively correlated with the presence of a wild-type p53 in the tumors. Additionally, the miRNA expression and p53 status partially correlated with tumor grade, as almost all cancer samples that contained a mutant p53 and expressed high levels of the miRs were derived from high-grade tumors.

We thus revealed a recurring signature of miRs that are coordinately regulated both in primary human cells *in vitro* and in human breast tumors *in vivo*. We suggest that these miRs are repressed by wild-type p53 during both normal growth and cancer progression.

The presented clusters contain families of paralogous cancer-related miRNAs

Interestingly, 15 miRs represented in the three clusters (Figure 1) are transcribed from three homologous genomic loci, reported earlier to be paralogs that evolved from a common evolutionary origin (Tanzer and Stadler, 2004). These include miRs-106b/93/25, which reside within an intron of the cell-cycle gene 'minichromosome maintenance protein 7' (*MCM7*); miRs-17/18a/19a/20a/19b-1/92a-1 (miR-17-92 polycistron), which are transcribed as the non-coding RNA *c13orf25*, and miRs-106a/18b/20b/19b-2/92-2 (miR-106a-92

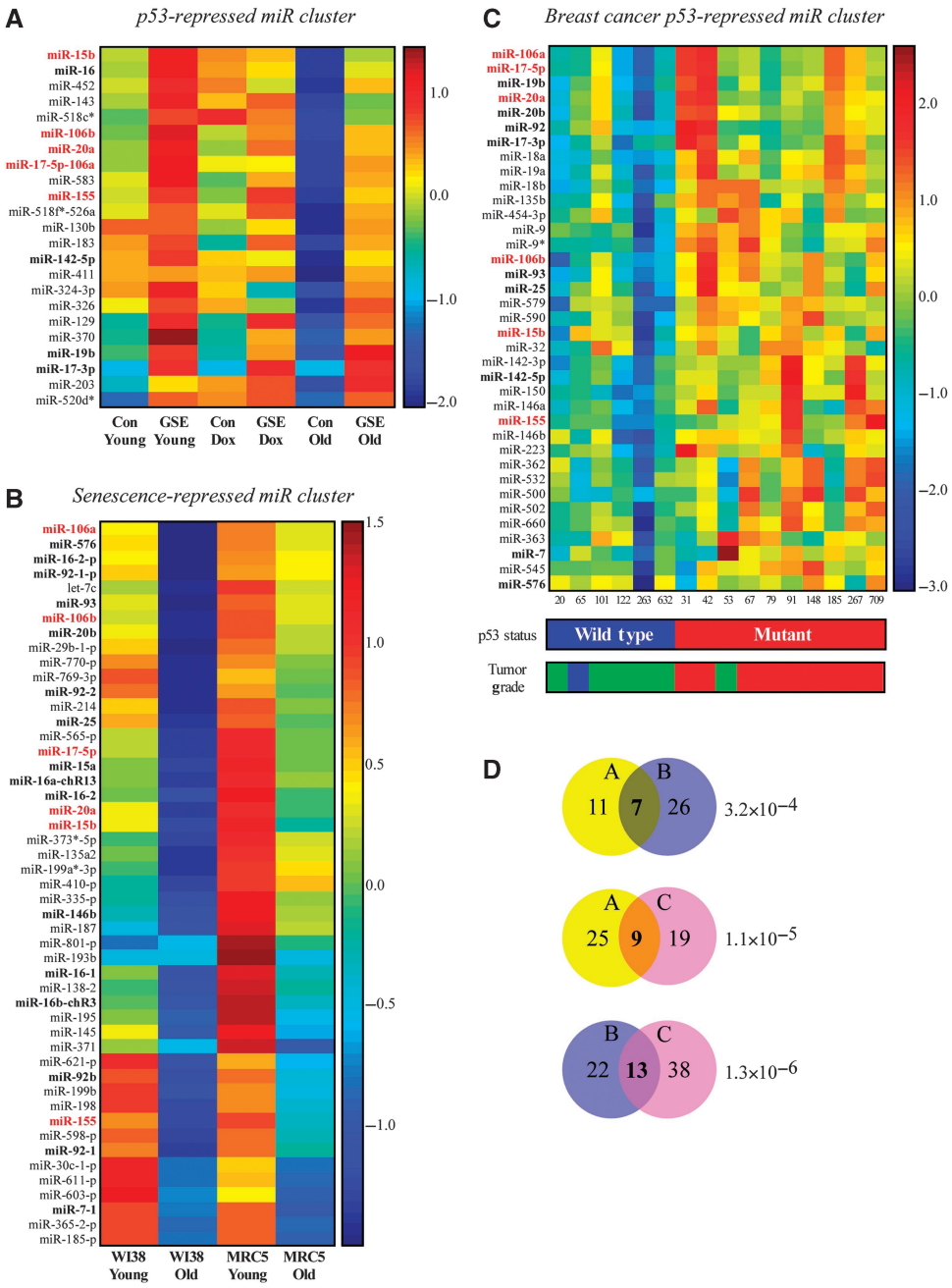


Figure 1 miRNA clusters derived from three different datasets. The figure depicts three different expression matrices for miRNA clusters that originated from three microarray experiments. miRs appearing in all three clusters are indicated in red bold font. miRs appearing in two clusters are indicated in black bold font. (**A**) The '*p53-repressed miR cluster*.' Primary WI-38 cells that were infected with the p53-inactivating peptide GSE56 (GSE) and their empty vector control counterparts (Con) were analyzed for miRNA expression at early passage (Young), after doxorubicin treatment (0.2 μg/ml, 24 h) of early passage cells (Dox), and at the onset of replicative senescence (Old). A cluster of miRs that were repressed by p53 at normal conditions and in senescent cells is presented. (**B**) The '*Senescence-repressed miR cluster*.' Primary WI-38 and MRC5 cells were analyzed for miRNA expression at early passage (Young) and at the onset of replicative senescence (Old). A cluster of miRs that were repressed upon senescence in both cell types is presented. (**C**) The '*breast cancer p53-repressed miR cluster*.' Human primary breast cancers described by Sorlie *et al* (2006) and Naume *et al* (2007) were analyzed for their miRNA profiles. A cluster of miRs, the expression of which was anticorrelated with the presence of a wild-type p53 in the tumor is presented. p53 status was determined using TTGE and sequencing of exons 2–11. Grading was performed using histopathological evaluation according to the modified Scarff–Bloom–Richardson method and is represented by blue for grade 1, green for grade 2 and red for grade 3. (**D**) Venn diagrams depicting the overlaps between cluster pairs. The values in each circle represent the number of miRs from the indicated cluster that was detected by the array corresponding to the second cluster. The values in the circle overlapping regions represent the number of miRs that are shared between the two clusters. Hypergeometric *P*-values on the size of the overlaps are provided.

polycistron), which are clustered on chromosome X. Our data indicate that not only were these miRNA sequences and genomic organization conserved during evolution but also was their transcriptional regulation. Additional well-represented miRs in the clusters include the miR-15b/16 polycistron and miR-155.

Many members of the clusters are overexpressed in various tumors, consistent with the frequent p53 loss of function in cancer, and some were shown to possess oncogenic functions. For example, miR-92, miR-106a, miR-17-5p, miR-20a and miR-155, which appear in at least two expression clusters, were reported to be overexpressed in solid tumors (Volinia *et al*, 2006). Members of the miR-17-92 polycistron are overexpressed in lymphomas and in lung and colorectal carcinomas (He *et al*, 2005; Schetter *et al*, 2008), and were shown to accelerate tumor growth (O'Donnell *et al*, 2005). Interestingly, the *MCM7* gene that contains three of the clusters' miRs in its intron (miRs-106b/93/25) is amplified or overexpressed in diverse types of cancers (Ren *et al*, 2006), as are its resident miRs (Petrocca *et al*, 2008).

Representative miRNAs show p53-dependent repression during senescence in many cell types

To further validate our data, we generated two additional human isogenic cell culture pairs from the IMR90 lung primary fibroblasts and from prostate-cancer-associated fibroblasts. Each culture was infected with a retrovirus encoding for either a small hairpin RNA (shRNA) targeting p53 (p53i) or a control shRNA (con), and grown until the onset of replicative senescence. p53 knockdown, which significantly reduced the mRNA and protein levels of both p53 and its target p21, delayed the onset of senescence by approximately 10 population doublings (PDLs) (Figure 2A; Supplementary Figure S2C). For these cell types, as well as for the WI-38 cells, we compared the levels of representative miRs that appear in the three expression clusters using TaqMan miRNA assays. Analyses of miR-106b and miR-17-5p, as well as their host transcripts *MCM7* and *c13orf25*, respectively, revealed transcriptional repression upon replicative senescence in all three tested cell cultures in a manner that was partially or completely p53 dependent (Figure 2B). Additionally, the non-coding RNA BIC and its resident miR-155 were also transcriptionally repressed in a p53-dependent manner upon replicative senescence (Supplementary Figure S2D).

Co-clustering of miRNA and mRNA expression data from human breast cancers reveals two distinct groups of p53-repressed miRNAs

To gain further insights into the regulation and function of the identified miRNA signature, we exploited previously published mRNA profiling (Sorlie *et al*, 2006; Naume *et al*, 2007), performed on the same set of breast cancer specimens from which the '*breast cancer p53-repressed miR cluster*' was derived. The mRNA and miRNA array data were combined into one set of expression profiles, and were clustered into 40 co-clusters; each may consist of both miRs and mRNAs (Figure 3). Interestingly, the members of the '*breast cancer p53-repessed miR cluster*' were separated into two distinct

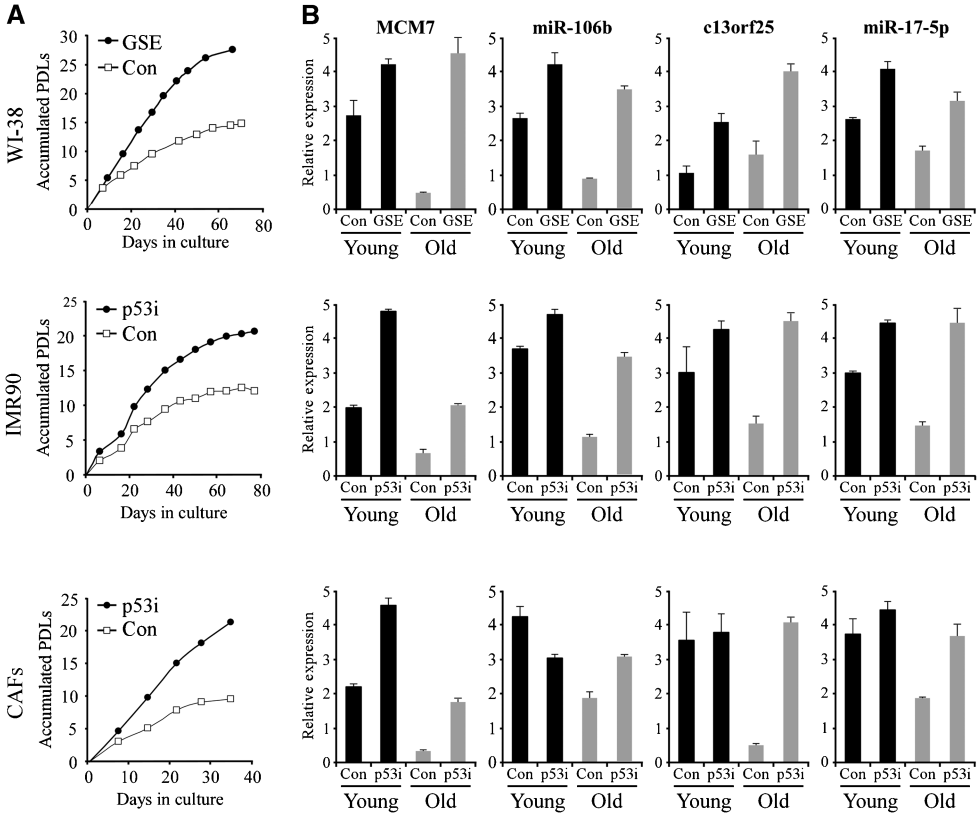


Figure 2 Validation of microarray data. Inactivation of p53 by the GSE56 peptide (GSE) or shRNA (p53i) in three different human primary fibroblasts delays replicative senescence and attenuates the repression of miRs and their hosts upon senescence. (**A**) Growth curves for the human primary fibroblasts WI-38 and IMR90 and for the prostate cancer-associated fibroblasts (CAFs) PF179. PDLs, population doublings. (**B**) QRT–PCR for miR-106b and miR-17-5p, and their host transcripts *MCM7* and *c13orf25*, respectively, in early passage (Young) versus late passage (Old) fibroblasts. Data are represented as mean ± s.d.

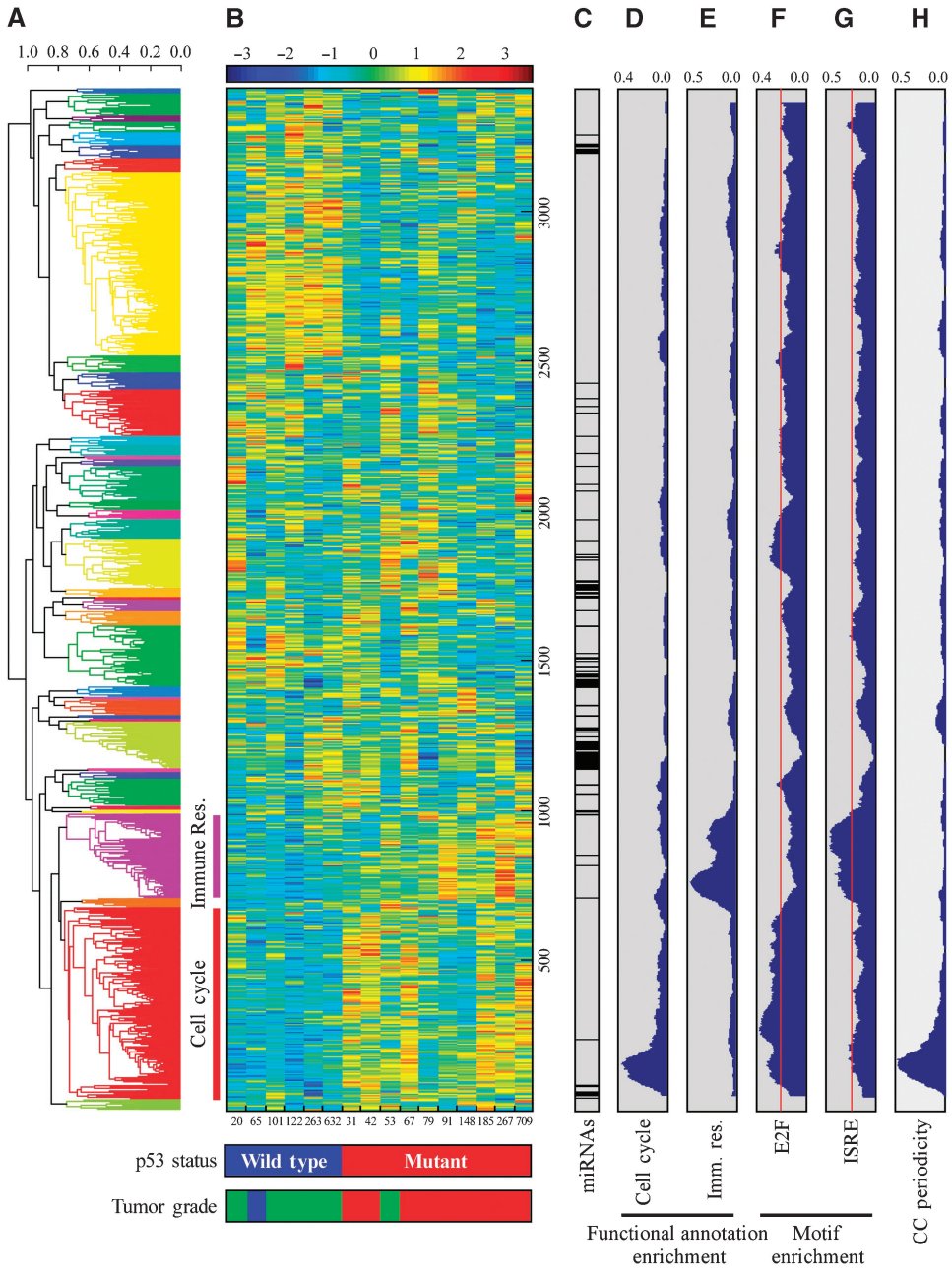


Figure 3 Co-clustering of miRNA and mRNA expression data from human breast cancers. (**A**) A dendrogram for the expression data based on hierarchical clustering and average linkage. Data were clustered into 40 clusters, which are indicated by different colors of the dendrogram. miRs from the '*breast cancer p53-repressed miR cluster*' were mapped to the red (*cell cycle*) and purple (*immune response*) clusters. (**B**) Expression matrix of the mRNAs and miRNAs analyzed. For p53 status and tumor grade analyses, see Figure 1 legend. Breast cancer samples are indicated by numbers below the matrix. (**C**) The bar indicates the position of miRs along the expression matrix. (**D, E**) Functional annotation analysis for '*cell cycle*' (C) and '*immune response*' (Imm. Res.) (D) terms. The plots represent the density (from 0 to 1) of mRNAs corresponding to each annotation term in windows of 100 genes. (**F**, **G**) Density plots for the appearance of the E2F and ISRE motifs, the most enriched elements in the *cell cycle* and *immune response* co-clusters, respectively. Red lines indicate the background levels of each motif, calculated as the fraction of genes in the genome containing the motif. (**H**) Density plot for cell-cycle periodic genes as defined by Whitfield *et al* (2002).

co-clusters with dissimilar characteristics, as revealed by functional annotation using DAVID (Dennis *et al*, 2003) and by promoter motif analyses using AMADEUS (Linhart *et al*, 2008) of the mRNAs from each co-cluster. Of 37 miRs, 31 were co-clustered (Figure 3A, red cluster; Supplementary Table S2) with genes highly enriched for cell-cycle-related annotations (Figure 3D; enrichment *P*-value for 'cell-cycle' annotation=8.5×10^{-20}, see Supplementary Table S3A for other enriched annotations) and for regulatory binding motifs of known cell-cycle-related TFs such as E2F (Figure 3F; *P*-value=2.5×10^{-16}). All members of the three paralogous polycistrons described above were included in this '*cell-cycle-associated co-cluster*,' as were their hosts *MCM7* and *c13orf25*. Supporting the notion that this cluster consists mainly of

cell-cycle-related genes, we compared the mRNA dataset to a previously published list of genes expressed in a cell-cycle periodic manner (Whitfield *et al*, 2002), and found a significant enrichment of periodically expressed genes in this cluster (Figure 3H; *P*-value=4.9×10^{-17}). Another co-cluster, which contained six miRs (Figure 3A, purple cluster; Supplementary Table S2), was comprised of genes highly enriched for immune response-related functions (Figure 3E; *P*-value for 'immune response' annotation=1.2×10^{-48}, see Supplementary Table S3B for other enriched annotations). The promoters of these genes were enriched for immune response-related motifs such as interferon-stimulated-responsive element (Figure 3G; *P*-value=1.62×10^{-12}) as well as for the IRF and NF-κB motifs (data not shown). Importantly, both mRNAs and miRNAs of both co-clusters were downregulated in tumors that harbor wild-type p53, suggesting that our miR signature belongs to a larger transcriptional program that mediates p53-dependent gene repression of both RNA types.

The miRNAs from the 'cell-cycle-associated co-cluster' are associated with p53 and E2F in a proliferation-related regulatory network

We have reported earlier the identification and characterization of an mRNA cluster termed the '*proliferation cluster*' that consists mainly of cell-cycle-related genes (Milyavsky *et al*, 2005). This cluster emerged from mRNA profiling of an *in vitro* transformation process, in which primary WI-38 cells were gradually transformed into tumorigenic cells. Importantly, the '*proliferation cluster*' is one of the most prominent expression signatures revealed when tumors are compared to normal tissues or when highly proliferating cells are compared to slow growing cells, and contains many cell-cycle periodic genes (Whitfield *et al*, 2006). The expression pattern of the '*proliferation cluster*' is highly similar to that of the '*p53-repressed miR cluster*'; i.e. the '*proliferation cluster*' mRNAs display p53-dependent downregulation. The similarity in expression patterns prompted us to hypothesize that both clusters share a common transcriptional program. It was shown earlier that the p53-mediated repression of the '*proliferation cluster*' was mediated through E2F (Tabach *et al*, 2005). Providing further support, the promoters of the mRNAs from the '*cell-cycle-associated co-cluster*' are highly enriched with E2F-binding motifs (Figure 3F), and in particular, a conserved E2F-binding site is found upstream of the three miRNA polycistrons (Supplementary Figure S3A and Supplementary Table S4). Confirming these predictions, chromatin immunoprecipitation analysis revealed that E2F1 binds to its conserved motifs upstream of each of the three polycistrons (Figure 4A). Similarly, it was shown recently that the miR-17-92 and the miR-106b/93/25 polycistrons are transcriptionally activated by E2F family members (O'Donnell *et al*, 2005; Sylvestre *et al*, 2007; Woods *et al*, 2007; Petrocca *et al*, 2008).

In view of the above, it appears conceivable that the miRs from the '*cell-cycle-associated co-cluster*' are transcriptionally activated by E2F, and that p53 exerts its repression through E2F inhibition. In agreement with the observed downregulation of the '*p53-repressed miR cluster*' in senescence, it was shown that E2F activity is significantly downregulated in senescent cells (Campisi and d'Adda di Fagagna, 2007). In addition, the miRs presented here are proposed to be novel members of the well-established '*proliferation cluster*.'

The p53-dependent repression of miRNAs from the 'cell-cycle-associated co-cluster' is mediated through E2F1

To experimentally test whether the cell-cycle-associated miRs and their host mRNAs are transcriptionally activated by E2F1, we infected primary WI-38 cells with E1A, a viral oncoprotein that disrupts pRb-E2F complexes and leads to an upregulation of the endogenous E2F activity (Fattaey *et al*, 1993). As expected, stable overexpression of E1A resulted in elevated levels of candidate miRNAs and host mRNAs, which were part of the '*cell-cycle-associated co-cluster*,' that together represent all three paralogous polycistrons (Figure 4B). Specifically, the *MCM7* gene and its resident miRNAs miRs-106b/93/25; the non-coding RNA *c13orf25* and its resident miR-17-5p; and miR-106a, which represents the miR-106a-92 polycistron, were all upregulated following E2F activation. We note that the level of miR-155, which belongs to the '*immune response co-cluster*', was not upregulated by E2F (Supplementary Figure S3E).

To investigate the kinetics of the miRNA's transcriptional activation by E2F1, we infected WI-38 cells with a retrovirus encoding for an E2F1 protein fused to a modified estrogen receptor (ER) ligand-binding domain. Treatment of ER-E2F1-expressing cells with 4-hydroxytamoxifen (4-OHT) permits ER-E2F1 translocation to the nucleus, thereby inducing its transactivation activity. As depicted in Figure 4C, as early as 4 h after ER-E2F1 activation by 4-OHT, significant upregulation of the candidate miRNAs and host transcripts was already evident. The miRNAs and their hosts peaked after 8–10 h of 4-OHT treatment, similarly to *Cyclin E*, a known E2F1 target. It is noted that upregulation of *MCM7* and its resident miRNAs following 4-OHT treatment was also observed in ER-E2F1 expressing lung carcinoma cells (H1299) and osteosarcoma cells (U2OS) (Supplementary Figure S3B and C). Finally, to strengthen the notion that E2F1 directly transactivates the miRNAs, we treated ER-E2F1 expressing WI-38 cells with 4-OHT in the presence or absence of cycloheximide, which inhibits protein biosynthesis and should attenuate the induction of the miRNAs if translation of a secondary mediator is required. As depicted in Supplementary Figure S3D, the induction of the miRNAs was not inhibited by cycloheximide. Altogether, these results indicate that E2F1 can directly bind its cognate sites upstream of the polycistronic miRNAs and activate their transcription.

Having shown that representative miRs are activated by E2F1 in our system, we set to test whether their p53-dependent repression is mediated through modulation of E2F1 activity. To that end, we infected WI-38 cells with a retrovirus encoding for either an shRNA targeting p53 (p53i) or a control shRNA (con) and treated them with Nutlin-3, a small molecule that stabilizes the p53 protein by inhibiting its Mdm2-dependent ubiquitylation and degradation (Vassilev *et al*, 2004), thereby inducing p53 activation in a non-genotoxic manner. Nutlin treatment resulted in a robust p53 protein accumulation, accompanied by p21 mRNA and protein induction (Figure 5A

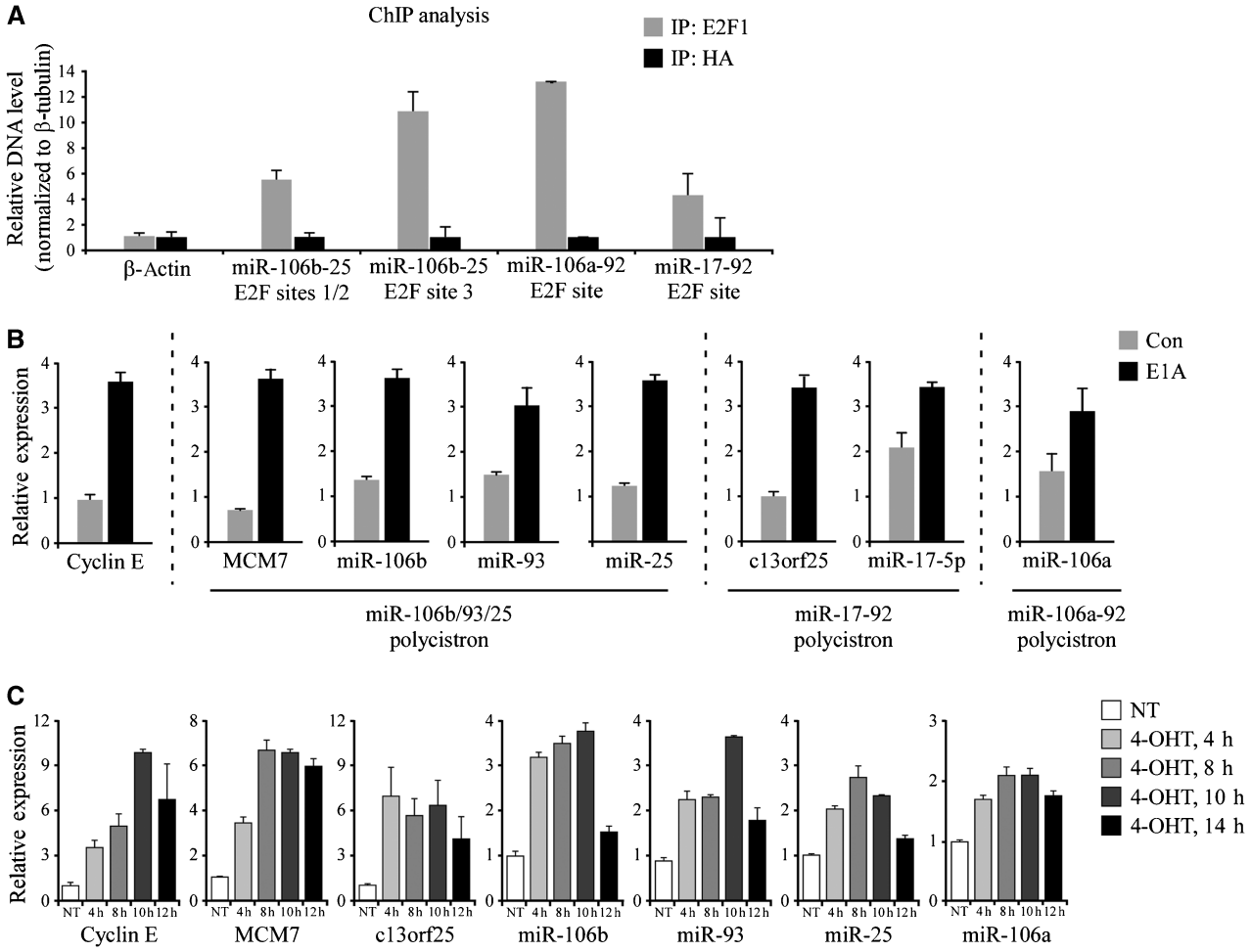


Figure 4 E2F binds the promoters of the paralogous miRNA polycistrons and induces their transcription. (**A**) E2F1 binds conserved E2F sites upstream of each of the paralogous miRNA polycistrons. ChIP analysis was performed on U2OS cells with an anti-E2F1 antibody (IP: E2F1) and a control antibody against HA (IP: HA). The precipitated DNA was measured using QRT–PCR. The β-actin gene serves as a negative control for E2F1 binding. Values were normalized to the levels of β-tubulin. For schematic representation of the polycistrons' genomic organization and corresponding E2F sites, see Supplementary Figure S3A. (**B**) E2F activation by E1A induces the polycistronic miRNAs. WI-38 cells were infected with the oncoprotein E1A or a control vector (Con) and selected with puromycin. QRT–PCR revealed upregulation of the known E2F1 target, *Cyclin E*, as well as of host transcripts and miRNAs, representatives of the three paralogous polycistrons (miRs-106b/93/25, miR-17-92 and miR-106a-92). (**C**) E2F1 activation results in rapid induction of the polycistronic miRNAs. WI-38 cells were stably infected with ER-E2F1 and treated with 4-OHT (300 nM) for the indicated time periods. QRT–PCR analysis was performed to measure the levels of miRNAs and mRNAs. QRT–PCR data are represented as mean ± s.d.

and B), which was completely abrogated in the p53i cells. Remarkably, E2F1 mRNA and protein levels were downregulated upon Nutlin treatment in a p53-dependent manner. *Cyclin E* showed a similar pattern, supporting the notion that E2F1 downregulation was accompanied by a reduction in E2F activity. Accordingly, both *MCM7* and its resident miR-106b were significantly downregulated in a p53-dependent manner (Figure 5A) along with other miRs from the '*cell-cycle-associated co-cluster*' (data not shown). Thus, treatment with Nutlin, a non-genotoxic p53 activator, resulted in a p53-dependent transcriptional repression of mRNAs and miRNAs with associated cell-cycle functions. To substantiate the causal relationship between the p53-dependent reduction of E2F1 activity and the repression of the miRs and their hosts, we treated control and E1A-expressing WI-38 cells with Nutlin. As depicted in Figure 5C and D, E1A induced the expression and prevented the Nutlin-dependent repression of E2F1 as well as of its target *Cyclin E*. Most significantly, E1A abolished the downregulation of *MCM7* and miR-106b upon Nutlin treatment. A similar pattern was observed for miR-17-5p and its host *c13orf25* (data not shown). Finally, we stably knocked down E2F1 using retroviral-encoded shRNA in WI-38 cells in combination with Nutlin treatment, and measured the levels of miRs-106b/25/93 (Figure 5E), as well as the protein levels of p53, p21 and E2F1 (Figure 5F). Indeed, the knock down of E2F1 resulted in reduced levels of the miRNAs. Supporting the notion that repression of the miR-106b/93/25 polycistron by p53 is mediated through E2F1 inhibition, Nutlin treatment of the E2F1-knockdown cells had very little effect as compared with the control cells.

We therefore conclude that E2F1 inhibition by p53 is necessary for the downregulation of *MCM7* and its resident miRNAs. The same mechanism may underlie the p53-dependent downregulation of additional miRs from the

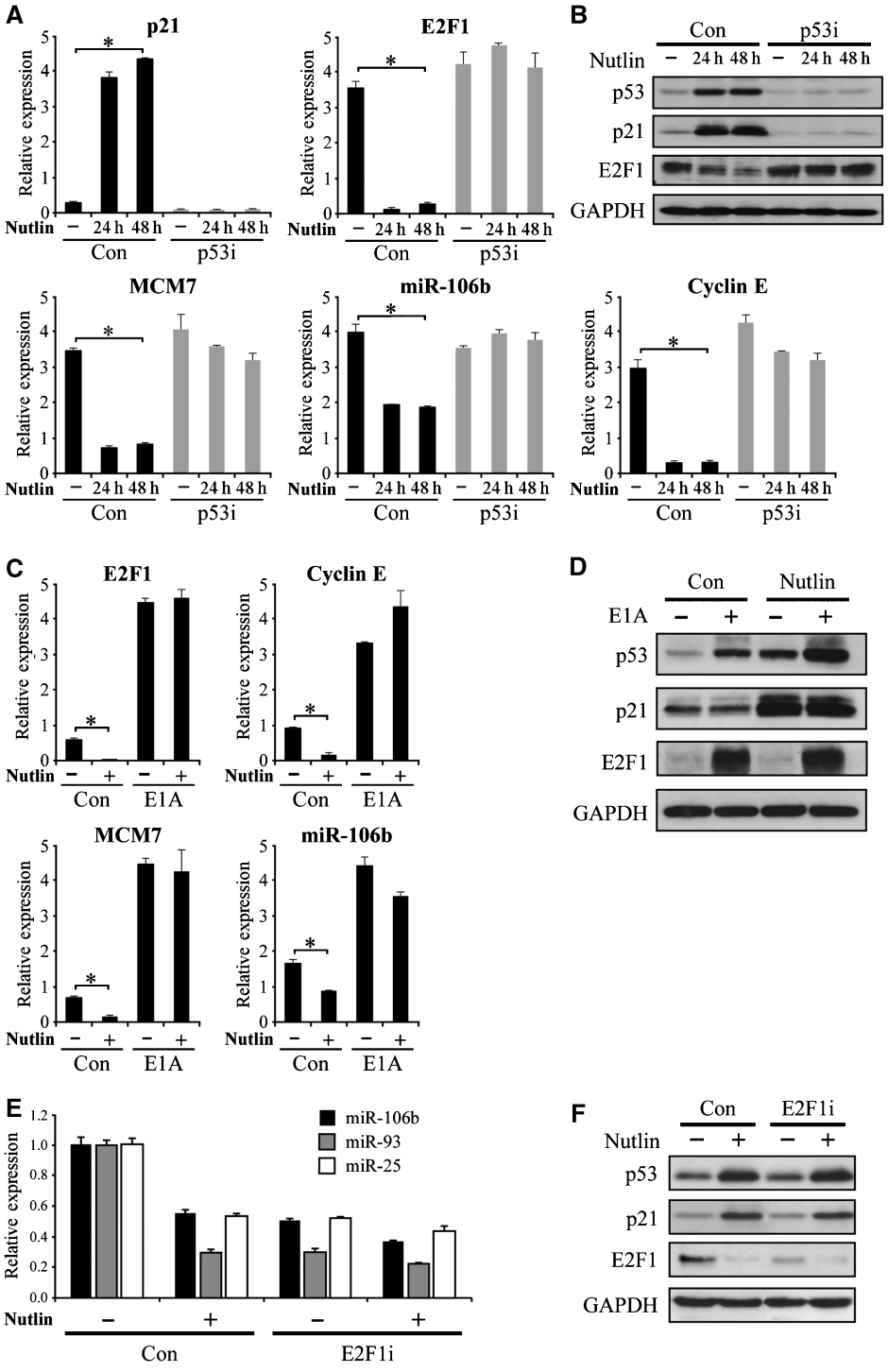


Figure 5 *MCM7* and miR-106b are repressed by Nutlin-activated p53 in an E2F-dependent manner. (**A**, **B**) WI-38 were infected with a retrovirus encoding for either a small hairpin RNA targeting p53 (p53i) or a control shRNA (Con) and treated with 10 μM Nutlin-3 for 24 or 48 h. QRT–PCR (A) and western blotting (B) analyses demonstrated p53 stabilization by Nutlin, which resulted in activation of *p21* and repression of E2F1 mRNA and protein levels. *MCM7* and its resident miR-106b were repressed in a p53-dependent manner upon Nutin-3 treatment. (**C**, **D**) WI-38 cells were infected with E1A or an empty vector control (Con) and treated with 10 μM Nutlin-3 for 24 h. E1A elevated E2F transactivation activity, resulting in the induction of *Cyclin E* and *E2F1* itself as well as of *MCM7* and miR-106b. Nutlin treatment of the control cells repressed transcription of E2F1 and its targets. E1A abolished this repression, indicating that the repression of E2F1 by p53 is necessary for the p53-dependent downregulation of *MCM7* and miR-106b. In (A, C), statistically significant difference in expression (*t*-test; *P*-value < 0.01) between the non-treated samples and the Nutlin-treated samples (at both 24 and 48 h) is marked by asterisks. (**E**, **F**) WI-38 cells were infected with a retrovirus encoding for either a small hairpin RNA targeting E2F1 (E2F1i) or a control shRNA (Con) and treated with 10 μM Nutlin-3 for 48 h. QRT–PCR (E) and western blot (F) analyses demonstrated repression of E2F1 and its targets miRs-106b/93/25, as well as activation of p53 and *p21* upon Nutlin treatment. E2F1 knockdown mimicked the effect of Nutlin treatment in repressing the miR-106b/93/25 polycistron. Nutlin treatment in the presence of E2F1 shRNA had little effect on the miRs, indicating that E2F1 inhibition mediates the repression of the miRs by Nutlin-activated p53. GAPDH protein levels serve as loading controls in (B, D, F). QRT–PCR data are represented as mean ± s.d.

'*cell-cycle co-cluster*' and, more specifically, the three paralogous polycistrons.

Cell-cycle-associated miRNAs target key cell-cycle regulators and affect pivotal characteristics of proliferation

Next, we set out to identify the functions of the p53-repressed miRs. We focused on the miR-106b/93/25 polycistron as a representative member of the large family of miRs that includes also the miR-17-92 and miR-106a-92 polycistrons. We overexpressed the genomic region encoding miRs-106b/93/25, which corresponds to an intron of the *MCM7* gene in young WI-38 cells and in MCF10A mammary epithelial cells, both characterized by low basal expression of these miRs. Following our previous computational prediction of E2F and miR-106b/93/25 involvement in a FFL, in which they both target a mutual set of genes (Shalgi *et al*, 2007), we compiled a list of their mutual predicted targets (Supplementary Table S5). Interestingly, many of these predicted targets participate in cell-cycle regulation (*P*-value for 'cell-cycle' annotation enrichment=1.4×10^{-10}). Another key cell-cycle regulator, p21, was recently reported as a target for miR-106b (Ivanovska *et al*, 2008), and is a known target of both E2F1 (Gartel *et al*, 1998) and p53. We then measured the protein levels of selected predicted targets in the miR-106b/93/25-overexpressing cells (Figure 6). We observed downregulation of *p21*, as well as of pRB and p130, which were suggested earlier, based on reporter assays, as potential targets of miR-106a and the miR-17-92

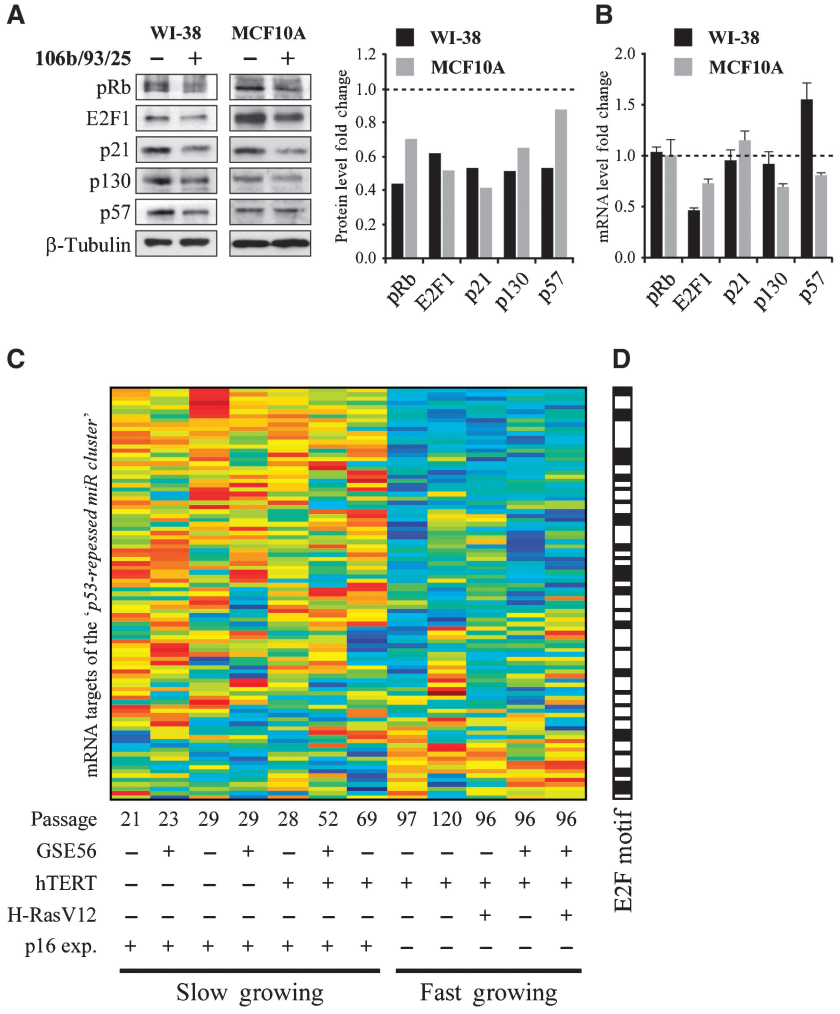


Figure 6 Overexpression of miR-106b/93/25 polycistron results in silencing of cell-cycle-related genes. WI-38 primary fibroblasts and MCF10A mammary cells were infected with a retrovirus encoding for either the genomic region that contains miRs-106b/93/25 or an empty vector control. (**A**) Western blot analysis of cell-cycle-regulating targets of the overexpressed miRs. Overexpression of miRs-106b/93/25 reduced the protein levels of E2F1, pRb, p130, E2F1 and p21 in both cell types and of p57 in WI-38 cells. β-Tubulin levels serve as a loading control. The scanned blots were analyzed using the ImageJ software. Values represent the fold change of each protein relative to the empty vector-infected cells, and were normalized to the levels of β-tubulin. (**B**) QRT–PCR analysis of the mRNA levels of the genes presented in (A). Values represent the fold change of each mRNA relative to the empty vector-infected cells. Data are represented as mean ± s.d. (**C**) Expression pattern of predicted targets of at least five miRs from the '*p53-repressed miR cluster*.' mRNA expression levels were derived from WI-38 cells that underwent immortalization and gradual *in vitro* transformation (the cell status is indicated below) as described by Milyavsky *et al* (2003). This expression pattern was found to be significantly coherent (EC score=0.14, EC *P*-value=5×10^{-3}). See Supplementary dataset S4 for the expression values. (**D**) Promoter analysis performed on the genes from (C) (using AMADEUS) revealed enrichment for E2F motif (*P*-value=2.2×10^{-13}). Genes with E2F motif in their promoter are indicated in black in the bar on the right (and in Supplementary dataset S4).

cluster, respectively (Volinia *et al*, 2006; Wang *et al*, 2008). Interestingly, E2F1, which was shown to be a target of miR-17-5p and miR-20a (O'Donnell *et al*, 2005), and is predicted by PicTar (Krek *et al*, 2005) to be a target for both miR-106b and miR-93, was significantly downregulated as well. We also observed downregulation of p57 in WI-38 cells, in agreement with PicTar predictions (Figure 6A; Supplementary Figure S4). Notably, these proteins have defined functions in the regulation of the cell cycle, most of them being negative regulators of proliferation. As the mRNA levels of p57, p21, pRb and p130 did not decrease in WI-38 cells and only marginally in MCF10A cells (Figure 6B), the reduction in their protein levels most likely stems from translational inhibition and not mRNA degradation. Considering the above, it is unlikely that the reduction in the targets' protein levels stems from reduced E2F transcriptional activity. In contrast, E2F1 mRNA levels were reduced in both cell lines that express the miR-106b/93/25 polycistron, in agreement with Petrocca *et al* (2008).

To gain independent support for the role of the '*p53-repressed miR cluster*,' we investigated the expression pattern of genes that are targeted by these miRs in the above-mentioned transformation system, where primary WI-38 cells were gradually transformed into tumorigenic cells (Milyavsky *et al*, 2005). Interestingly, targets harboring predicted sites for multiple miRs from the '*p53-repressed miR cluster*' within their 3′-UTR had significantly coherent expression patterns (Pilpel *et al*, 2001) during the transformation process (Figure 6C). This observation means that genes that are targeted by multiple miRs from the '*p53-repressed miR cluster*' are significantly co-expressed during the transformation process as compared with random sets of genes. Furthermore, not only are these targets expressed similarly to one another but also the actual expression pattern of many of them is consistent with the pro-proliferative role of the miRs that regulate them, i.e. the expression of the majority of the genes in this target set was decreased when cells gained the accelerated proliferation phenotype (designated as 'fast growing'). Furthermore, promoter analysis, using the AMADEUS algorithm (Linhart *et al*, 2008), of these predicted target genes, revealed the E2F-binding site as one of the most highly enriched motifs (P-value$=2.2\times10^{-13}$). This supports our general notion that E2F cooperates with the '*p53-repressed miR cluster*' in regulating shared targets at the transcriptional and post-transcriptional levels. This suggests that the FFL consisting of E2F and miRs-106b/93/25 may be deregulated during cancer progression. We note that a fifth of the predicted targets of the miR cluster show a very different expression pattern (Figure 6C, bottom part), which may indicate more complex regulatory interactions.

Having shown the molecular effects of the overexpression of the miR-106b/93/25 polycistron, we tested whether proliferation-related parameters such as growth rate, colony formation efficiency (CFE) and replicative senescence are affected by these miRs. As these miRs are significantly repressed by p53 during senescence, and considering the fact that they target several antiproliferation regulators, we predicted that their overexpression, similarly to p53 inactivation, would accelerate cellular growth rate and delay senescence. Indeed, as depicted in Figure 7, the miR-106b/93/25-overexpressing WI-38 cells demonstrated a moderate acceleration in proliferation rate and an increased fraction of S-phase cells (24% compared to 18%). Strikingly, these cells displayed a pronounced increase in the efficiency of young cells to form colonies when seeded at low density and reduced senescence-associated beta-galactosidase (SA-β-Gal) staining at late passages, indicating a delay of replicative senescence. Additionally, we evaluated the effect of miR-106b/93/25 overexpression in WI-38 cells on their CFE in combination with p53 inactivation. As depicted in Supplementary Figure S5, miR-106b/93/25 enhanced the CFE of both active and inactive p53-expressing cells. However, the effect of the overexpressed miRs was much more pronounced in the active p53 cells, augmenting their CFE by 20-fold as compared with only 2.6-fold increase in CFE in the inactive p53 cells. In fact, the effect of overexpression of the miRs on the CFE of the control cells was comparable to that of p53 inactivation. These observed phenotypes suggest that the transcriptional repression of miR-106b/93/25 and their paralogs mediates part of the antiproliferative effects of p53.

Discussion

In the present study, we elucidate a complex regulatory network involving a group of cancer-related miRs. In this network, E2F1 transcriptionally controls the miR-106b/93/25 polycistron and its paralogs, and together they regulate a mutual set of target genes. In concordance with the growth acceleration that resulted from the overexpression of these miRs, many of their targets are considered antiproliferative cell-cycle regulators. Importantly, this intricate FFL is repressed by p53 through inhibition of E2F1. A schematic model for the proposed network is presented in Figure 8.

Employing three independent experiments, we identified a novel miR signature that is transcriptionally repressed by p53 in human primary cells and in breast cancers. Consistent with p53 function, many signature members, including the three paralogous polycistrons and miR-155, are considered oncogenic miRs and are overexpressed in diverse types of tumors (Eis *et al*, 2005; He *et al*, 2005; Volinia *et al*, 2006; Yanaihara *et al*, 2006). miR-15b and miR-16, which are considered to be tumor suppressor miRs (Cimmino *et al*, 2005), are exceptions in this regard.

Upon diverse stress stimuli, p53 is known to regulate different subsets of genes, resulting in alternative cellular outcomes (Oren, 2003). Consistently, the repression of the miRs was restricted to non-genotoxic contexts, namely, replicative senescence and Nutlin-induced Mdm2 inhibition, as doxorubicin treatment did not result in transcriptional repression of the miR cluster despite p53 activation. Our earlier studies have shown that Nutlin treatment induces p53-dependent senescence accompanied by upregulation of miR-34 (Kumamoto *et al*, 2008). Therefore, the repression of the miRs might be specific to p53-induced senescence triggered in this study by cellular aging or Nutlin treatment. As p53 is capable of inducing senescence *in vivo* (Xue *et al*, 2007), the observed miRNA repression in the wild-type p53-harboring breast tumors may be associated with increased senescence in these samples.

Co-clustering of coding mRNAs and microRNAs from the breast cancer study separated the p53-repressed miRs into two

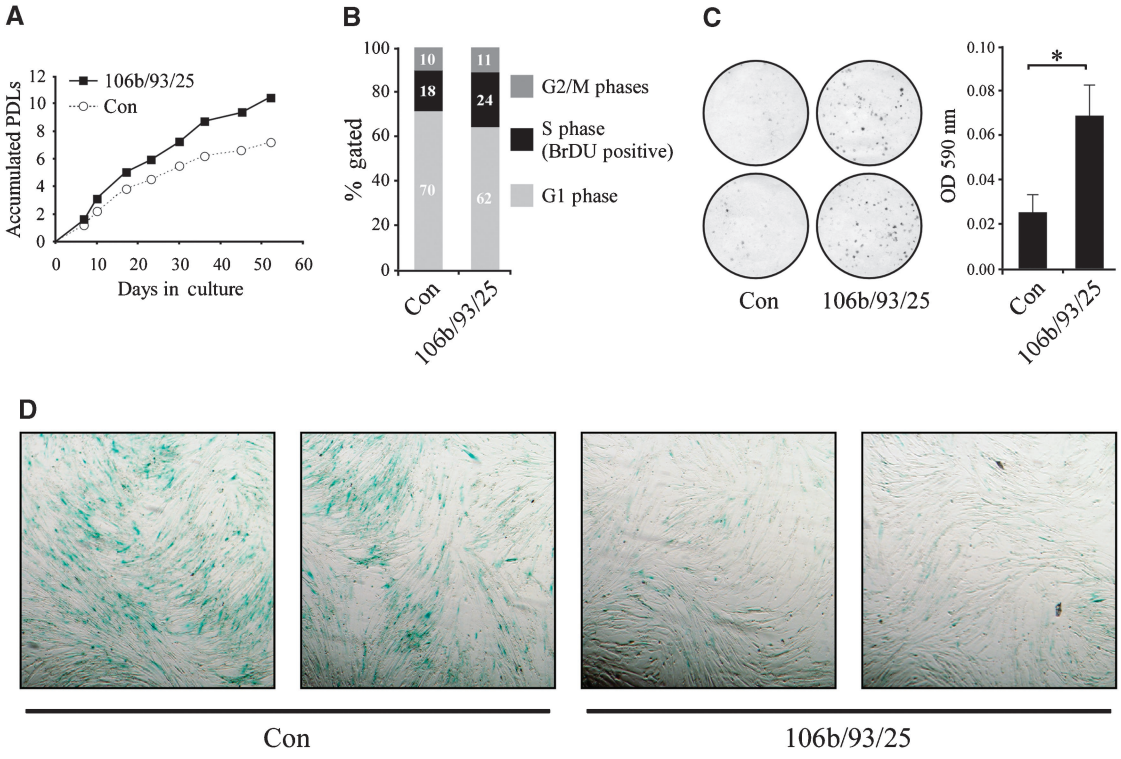


Figure 7 Overexpression of miR-106b/93/25 polycistron in WI-38 cells promotes proliferation. Overexpression of miR-106b/93/25 polycistron in WI-38 cells promotes proliferation. (**A**) Growth curves for control (empty vector) and miR-106b/93/25-overexpressing cells. PDLs, population doublings. The difference between the growth curves was analyzed by paired *t*-test of the number of PDLs in each passage, and was found to be statistically significant (*P*-value$=1.2\times10^{-4}$) (**B**) Cell-cycle analysis of BrDU-labeled cells using fluorescence cytometer. miR-106b/93/25-expressing cells demonstrated increased proportion of S phase (BrDU positive) cells. (**C**) Colony formation assay. Cells were plated at low density and grown for 2 weeks. Plates were stained with crystal violet (left). The crystal violet was extracted with acetic acid and quantified with a spectrophotometer using a 590 nm filter (right). The difference was statistically significant (*P*-value <0.05). (**D**) Senescence-associated β-galactosidase staining depicting decreased level of senescence in miR-106b/93/25-expressing cells as compared with their empty vector control counterparts (Con).

functional and regulatory categories, namely '*cell cycle*' and '*immune response*' (Figure 3). The expression-based division perfectly mirrored earlier reported functions of these miRs, including members of the three paralogous polycistrons that were co-clustered with cell-cycle-associated mRNAs, and were shown here and in additional reports (Hossain *et al*, 2006; Lu *et al*, 2007; Ivanovska *et al*, 2008) to promote cell proliferation. Another p53-repressed polycistron, miR-15b/16, was clustered with the cell-cycle genes. Indeed, members of this polycistron have been implicated in the regulation of cell-cycle progression (Linsley *et al*, 2007). The '*immune response-associated co-cluster*' included miR-155, which is commonly overexpressed in lymphomas (Eis *et al*, 2005), participates in the germinal center response (Thai *et al*, 2007), and is upregulated in chronic gastritis (Petrocca *et al*, 2008). The remaining members of this co-cluster were also implicated in immune responses, including miR-150 and miR-146 (Lodish *et al*, 2008) as well as miR-142 (Wu *et al*, 2007). Overexpression of miR-155 in primary cells did not affect the rate of proliferation (Supplementary Figure S3F), suggesting distinct functions for the members of the '*immune response-associated co-cluster*.' Interestingly, in addition to many cell-cycle-related coding genes reported to be repressed by p53, recent evidence indicates p53-mediated repression of immune response-related genes (e.g. interleukin-1β, interleukin-6 and *Cxcl1* (Buganim *et al*, submitted) and SDF-1 (Moskovits *et al*, 2006)).

For the first time, we demonstrate that the three paralogous polycistronic miRNAs are coordinately activated by E2F1. Importantly, we establish E2F1 as the mediator of the p53-dependent repression of miRs-106b/93/25 and suggest that this mechanism underlies the repression of the two additional paralogous polycistrons. Upon Nutlin treatment, E2F1 protein levels were dramatically downregulated in a p53-dependent manner. A similar phenomenon was described earlier, and was attributed to enhanced ubiquitylation of E2F1 by an unknown ligase, resulting in proteasome-mediated degradation (Ambrosini *et al*, 2007). However, we demonstrate that Nutlin treatment also results in a robust p53-dependent E2F1 mRNA repression, in agreement with an earlier observation that overexpression of p53, as well as p21, results in down-regulation of E2F1 mRNA (Ookawa *et al*, 2001). It is plausible that p53-mediated reduction in E2F1 protein inhibits E2F1 transcription as this gene contains an E2F motif in its promoter and is itself an E2F target gene (Johnson *et al*, 1994). Thus, even a slight reduction in E2F1 protein level might trigger a feedback loop that will result in significant reduction of both protein and mRNA levels. Considering this feedback loop, inhibition of E2F1 activity could also explain the observed repression of its mRNA and protein levels. Such inactivation may be indirectly mediated by the p53 target gene p21 through the inhibition of CDKs that inactivate the pocket proteins, which in turn inhibit E2F activity. Another mechanism for

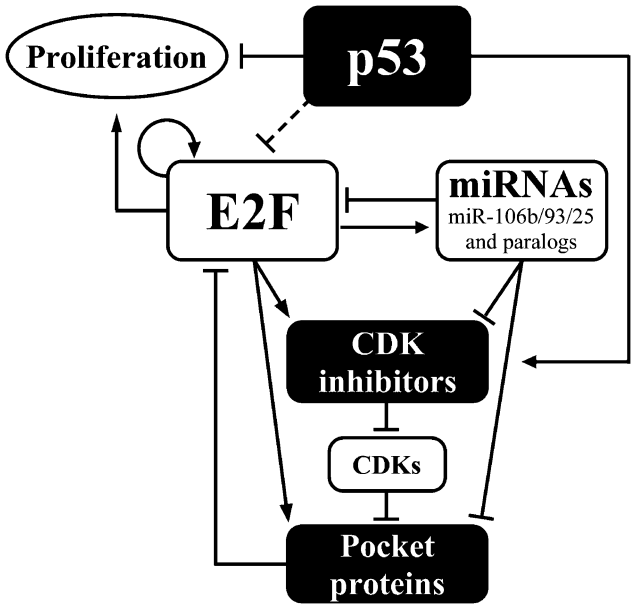


Figure 8 A schematic model for the cell-cycle regulatory network comprising E2F, p53, miRs and other cell-cycle regulators. Arrows correspond to direct transcriptional activation, whereas bar-headed lines represent direct or indirect inhibition mediated by the following mechanisms: post-transcription gene silencing (miRNAs and their targets), protein binding and inactivation (pocket proteins and E2Fs; as well as CDK inhibitors and CDKs, that in turn inhibit pocket proteins by phosphorylation). The circular arrow represents E2F self-activating ability. Possible mechanisms underlying the repression of E2F by p53 are detailed in the discussion.

E2F1 inactivation by p53 was recently suggested by identifying *BTG3* as a p53 target gene that directly binds E2F1 and inhibits its activity (Ou *et al*, 2007). Yet another possible mediator for p53-dependent E2F1 suppression is miR-34a, a direct transcriptional target of p53, which was recently suggested to induce senescence and to repress the E2F pathway (Tazawa *et al*, 2007).

Earlier studies have reported the effect of a single miR on a single target gene, such as miR-106b effect on p21 (Ivanovska *et al*, 2008). Others have described E2F-dependent activation of a single polycistron, miR-17-92 (Sylvestre *et al*, 2007; Woods *et al*, 2007), and miRs-106b/93/25 (Petrocca *et al*, 2008). We suggest that the three polycistrons, consisting of a family of 15 different miRs, are in fact transcriptionally co-regulated directly by E2F and indirectly by p53. In addition, many miRs in this family share highly similar seed sequences (Tanzer and Stadler, 2004). Thus, when an entire miR family is coordinately activated, its combinatorial and cumulative effects on mRNA targets may be profound. To recapitulate natural conditions, we combinatorially expressed three of the family miRs (miRs-106b/93/25), which are naturally co-transcribed, and demonstrated their effect on a set of target proteins. Future technologies allowing combinatorial knock down of an entire miR family may further establish their effects on other target genes. We show that miRs-106b/93/25 silence key members of the E2F pathway, including negative regulators of proliferation such as the pocket proteins pRb and p130 and the CDK inhibitors p21 and p57 (Figure 6). These and many other cell-cycle regulators are known E2F targets and are predicted to be silenced by miRs-106b/93/25 (Supplementary Table S5). Thus, we provide experimental evidence for our recent *in silico* predicted FFL motif (Shalgi *et al*, 2007).

We demonstrate here (Figure 6) a cancer-related manifestation of the concept of miR target avoidance (Farh *et al*, 2005; Stark *et al*, 2005). These studies introduced the concept of 'spatio-temporal avoidance,' showing that miRs and their targets tend to avoid being expressed in the same tissue or at the same developmental time, thereby assisting to determine differentiation boundaries and transitions. This avoidance could reflect a direct negative regulatory effect of the miR on its target. Alternatively, miR target avoidance may be mediated through a common transcriptional program controlling both the miRs and their targets.

The overexpression of miRs-106b/93/25 phenotypically mimicked p53 inactivation in WI-38 cells, as evident from an elevated rate of proliferation, increased CFE and delay of senescence. Importantly, induction of senescence, which we suggest to be partially mediated by the repression of the polycistronic miRs described above, is considered one of the main mechanisms by which p53 suppresses tumor formation (Xue *et al*, 2007).

In summary, we present here another arm of p53's tight control of cell proliferation, senescence and tumor suppression. This involves an elaborate network encompassing miRs and their targets, which modulate cell fate both during normal growth and in cellular senescence.

Materials and methods

Cell culture

WI-38, MRC5, IMR90 (obtained from the ATCC) and PFCA179 (provided by Dr H Klocker) cells were cultured in MEM with 10% FCS (fetal calf serum), 1 mM sodium pyruvate, 2 mM L-glutamine and antibiotics. U2OS and H1299 cell lines were cultured in DMEM and RPMI, respectively, with 10% FCS and antibiotics. MCF10A cells were maintained in DMEM F12 supplemented with 5% horse serum, 0.5 μg/ml hydrocortisone, 0.1 mg/ml insulin, 0.1 μg/ml cholera toxin and 10 ng/ml EGF. All cells were maintained in a humidified incubator at 37°C and 5% CO_2. Primary fibroblasts were passaged every 5–6 days. PDLs were calculated using the formula: PDLs=log(cell output/cell input)/log2. For colony formation assays, cells were plated at low density (0.1–0.2 cells/mm^2), grown for 10–14 days and stained with crystal violet.

Plasmids and retroviral infections

The retrovirus encoding for GSE56 was described by Milyavsky *et al* (2005). shRNAs targeting p53 (p53i), or mouse NOXA (control shRNA) were stably expressed using pRetroSuper and were described by Berkovich and Ginsberg (2003). pRetroSuper-E2F1 was described by Korotayev *et al* (2008). ER-E2F1 was described by Vigo *et al* (1999). E1A was expressed from pBabe-puro-E1A12S (a gift from K Helin). For expression of miRs-106b/93/25, a 1-kb human genomic fragment was cloned with the primers 5′-ggatcctatcctgcgcctttcc-3′ and 5′-cacatggccacagaagac-3′ into miR-Vec (Voorhoeve *et al*, 2006). For expression of miR-155, a 238-bp human genomic fragment was cloned with the primers 5′-gtggcacaaaccaggaag-3′ and 5′-tatccagcagggtgactc-3′ into miR-Vec. Retrovirus infection procedures were described by Milyavsky *et al* (2003).

RNA preparation and quantitative real-time PCR

RNA was extracted with TRI reagent (Molecular Research Center Inc.). For mRNA quantification, a 2 μg aliquot of total RNA was reverse transcribed using Bio-RT (Biolab) and random hexamers. Quantitative

real-time PCR (QRT–PCR) was performed using Platinum SYBR Green qPCR SuperMix (Invitrogen). mRNA levels were normalized to the level of GAPDH in the same sample. Primer sequences are listed in Supplementary Table S6A. For miRNA quantification, TaqMan miRNA assays (Applied Biosystems) were used according to the manufacturer's protocol. Levels were normalized to the U6 control gene. All QRT–PCR reactions were performed on ABI7300 machine. Results are presented as mean and standard deviation for two or three duplicate runs.

miRNA microarrays, data analysis and clustering

The miRNA profiling presented in Figure 1A was performed as follows: RNA was extracted from WI-38 cells using TRI reagent as described above, labeled with Hy5 and hybridized on Exiqon's miRCURY™ LNA Array (v.8.1) with a common reference Hy3-labeled RNA pool. Data are provided as Supplementary dataset S1. Two biological replicates were performed for each sample type. Hy5/Hy3 ratios were log2 transformed and filtered such that miRs that were undetected in 11 or 12 samples were discarded. Duplicates were averaged, such that each miR was represented by six values, corresponding to the six different samples. For each miR, a credibility value was calculated as one minus the average of the six standard deviations (s.d.) between the duplicates. A duplicate that had one missing value was set as the detected value and was assigned with high s.d. The 5% most non-credible miRs were discarded. Data were clustered using hierarchical clustering (average linkage), with 20 clusters. The miRNA profiling presented in Figure 1B was performed as follows: RNA was extracted from young and old WI-38 and MRC5 cells using TRIZOL (Invitrogen) according to the manufacturer's protocol and was used for biotin labeling and hybridization on the version 4.0 miR arrays (Ohio State University) as described by Liu *et al* (2004). Data were normalized and log2-transformed. Data are provided as Supplementary dataset S2. Data were clustered using hierarchical clustering (average linkage), with 10 clusters. The miRNA profiling presented in Figure 1C was performed as follows: RNA was extracted as described by Sorlie *et al* (2006). Samples were hybridized on Agilent's miRNA arrays (beta version of V1) at the Agilent's facilities in St Clara, US by HJ. Data are provided as Supplementary dataset S3. Two outlier samples (both belonging to the mutant p53 set) were discarded (see Supplementary Figure S6). Data were clustered using hierarchical clustering (average linkage), with 25 clusters.

Microarray data can be downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). Accession numbers are as follows: GSE12450 (the WI-38 p53-dependent senescence study), GSE12821 (the WI-38 and MRC5 senescence study) and GSE12848 (The Breast Cancer Study).

Co-clustering of miRNA and mRNA expression data from human breast cancer samples

mRNA array data were filtered and normalized as in Sorlie *et al* (2006) (GEO accession number GSE3155) and only samples that were included in the miRNA analysis were used. Data were log2-transformed, and replicate mRNA probes were averaged. Only variable mRNAs that differ by at least 1.5-fold from their median expression in at least 40% of the samples were considered. miRNA data were centered such that the intensity values for each miR were divided by their mean, and log2 transformed. The combined mRNA and miRNA data were then clustered using hierarchical clustering (average linkage) with 40 clusters. miRs from the '*breast cancer p53-repressed miR cluster*' were mapped to the resulting co-clusters, and were found to reside in two clusters, that were then analyzed for functional enrichment using DAVID (Dennis *et al*, 2003), and for enrichment of sequence motifs in their corresponding ENSMBL gene promoters using AMADEUS (Linhart *et al*, 2008).

Analysis of miRNA targets expression coherence

The entire set of miRNA expression profiles was clustered into 20 miR clusters based on the above expression data (WI-38 young versus senescent, along with p53 inactivation). Then, we compiled a set of predicted targets for the miRs from each cluster using PicTar (Krek *et al*, 2005). Specifically, for each of the 20 miR clusters, a series of potential sets of targets were created. The first set consisted of mRNAs predicted to be targeted by at least one miR from the miR cluster. The second set consisted of mRNAs predicted to be targeted by at least two miRs from the miR cluster, and so on. The expression coherence (EC) score, a measure of expression similarity (Pilpel *et al*, 2001), was then computed for each set of targets according to their expression described by Milyavsky *et al* (2005) (expression data are also available as Supplementary dataset S5). The most significantly coherent expression pattern belonged to the set of genes that had target sites for at least five miRs from the '*p53-repressed miR cluster*' (EC P-value$=5\times10^{-3}$). The expression values for this gene set in the data from Milyavsky *et al* (2005) and prediction of E2F sites are found in Supplementary Figure S4.

Immunoblot analysis

Western blots were performed as described by Milyavsky *et al* (2005). The following antibodies were used: α-p53 pAb H-47 (produced in our laboratory), α-p21 sc-377 (Santa Cruz), α-E2F1 sc-193 (Santa Cruz), α-GAPDH MAB374 (Chemicon), α-p130 sc-317 (Santa Cruz), α-p57 sc-8298 (Santa Cruz), α-pRb 554136 (Pharmingen), and α–β-tubulin T7816 (Sigma).

Cell-cycle analysis

Cells were labeled for 30 min with 10 μM BrdU (Sigma), fixed with 70% EtOH/HBSS (2 h, −20°C), treated with 2 M HCl/0.5% Triton, washed and treated with 0.1 M $Na_2B_4O_7$ pH 8.5, and stained with FITC-conjugated anti-BrdU (Becton Dickinson) and 10 μg/ml propidium iodide. Samples were analyzed using a FACSort machine (Becton Dickinson). At least 1×10^4 events were recorded per sample.

SA-β-Gal activity assay

Cells were fixed with 3% formaldehyde/PBS for 5 min, washed with PBS and incubated for 16 h at 37°C with a solution containing 1 mg/ml X-gal, 40 mM citric acid, sodium phosphate, pH 6.0, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 150 mM NaCl and 2 mM $MgCl_2$.

Chromatin immunoprecipitation

DNA–protein complexes were immunoprecipitated from U2OS cells using the ChIP assay kit (Upstate Biotechnology), according to the manufacturer's protocol with the following polyclonal antibodies: α-E2F1 sc-193 (Santa Cruz) and α-HA sc-805 (Santa Cruz); the latter served as a control for nonspecific DNA binding. The precipitated DNA was subjected to QRT–PCR analysis using specific primers corresponding to each predicted E2F site, as well as primers for normalization (β-tubulin) and negative control for E2F1 binding (β-actin coding region). Primer sequences are listed in Supplementary Table S6B.

Supplementary information

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

Acknowledgements

This study was supported by a Center of Excellence grant from the Flight Attendant Medical Research Institute; EC FP6 funding (contract no. 502983); FP7 funding (ONCOMIRS, agreement 201102); the Ben May Charitable Trust; the Israel Science Foundation and the Yad Abraham Center for Cancer Diagnosis and Therapy. This research was supported in part by the Intramural Research Program, NIH, NCI and

CCR. VR is the incumbent of the Norman and Helen Asher Professorial Chair Cancer Research at the Weizmann institute. YP is an incumbent of the Rothstein Career Development Chair in Genetic Diseases. RS is a fellow of the Horowitz Foundation for Complexity Sciences. GHN was supported by the Howard Hughes Medical Institute Grant for Graduate Medical Education. This publication reflects only authors' views. The European Commission is not liable for any use that may be made of the information herein. We thank Agilent Technologies for access to the early version of their miRNA arrays and for their support.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ambrosini G, Sambol EB, Carvajal D, Vassilev LT, Singer S, Schwartz GK (2007) Mouse double minute antagonist Nutlin-3a enhances chemotherapy-induced apoptosis in cancer cells with mutant p53 by activating E2F1. *Oncogene* **26:** 3473–3481

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297

Berkovich E, Ginsberg D (2003) ATM is a target for positive regulation by E2F-1. *Oncogene* **22:** 161–167

Campisi J, d'Adda di Fagagna F (2007) Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol* **8:** 729–740

Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH, Feldmann G, Yamakuchi M, Ferlito M, Lowenstein CJ, Arking DE, Beer MA, Maitra A, Mendell JT (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell* **26:** 745–752

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M, Croce CM (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci USA* **102:** 13944–13949

Dennis Jr G., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4:** P3

Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, Lund E, Dahlberg JE (2005) Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc Natl Acad Sci USA* **102:** 3627–3632

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

Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* **310:** 1817–1821

Fattaey AR, Harlow E, Helin K (1993) Independent regions of adenovirus E1A are required for binding to and dissociation of E2F–protein complexes. *Mol Cell Biol* **13:** 7267–7277

Gartel AL, Goufman E, Tevosian SG, Shih H, Yee AS, Tyner AL (1998) Activation and repression of p21(WAF1/CIP1) transcription by RB binding proteins. *Oncogene* **17:** 3463–3469

Ginsberg D, Mechta F, Yaniv M, Oren M (1991) Wild-type p53 can down-modulate the activity of various promoters. *Proc Natl Acad Sci USA* **88:** 9979–9983

He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, Jackson AL, Linsley PS, Chen C, Lowe SW, Cleary MA, Hannon GJ (2007) A microRNA component of the p53 tumour suppressor network. *Nature* **447:** 1130–1134

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM (2005) A microRNA polycistron as a potential human oncogene. *Nature* **435:** 828–833

Ho J, Benchimol S (2003) Transcriptional repression mediated by the p53 tumour suppressor. *Cell Death Differ* **10:** 404–408

Hossain A, Kuo MT, Saunders GF (2006) Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. *Mol Cell Biol* **26:** 8191–8201

Hussain SP, Harris CC (1999) p53 mutation spectrum and load: the generation of hypotheses linking the exposure of endogenous or exogenous carcinogens to human cancer. *Mutat Res* **428:** 23–32

Ivanovska I, Ball AS, Diaz RL, Magnus JF, Kibukawa M, Schelter JM, Kobayashi SV, Lim L, Burchard J, Jackson AL, Linsley PS, Cleary MA (2008) MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol Cell Biol* **28:** 2167–2174

Johnson DG, Ohtani K, Nevins JR (1994) Autoregulatory control of E2F1 expression in response to positive and negative regulators of cell cycle progression. *Genes Dev* **8:** 1514–1525

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ (2005) RAS is regulated by the let-7 microRNA family. *Cell* **120:** 635–647

Kern SE, Kinzler KW, Baker SJ, Nigro JM, Rotter V, Levine AJ, Friedman P, Prives C, Vogelstein B (1991) Mutant p53 proteins bind DNA abnormally *in vitro*. *Oncogene* **6:** 131–136

Korotayev K, Chaussepied M, Ginsberg D (2008) ERK activation is regulated by E2F1 and is essential for E2F1-induced S phase entry. *Cell Signal* **20:** 1221–1226

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N (2005) Combinatorial microRNA target predictions. *Nat Genet* **37:** 495–500

Kumamoto K, Spillare ES, Fujita K, Horikawa I, Yamashita T, Appella E, Nagashima M, Takenoshita S, Yokota J, Harris CC (2008) Nutlin-3a activates the p53 tumor suppressor to both down-regulate ING2 and up-regulate mir-34a, b and c expression and induce senescence. *Cancer Res* **68:** 3193–3203

Linhart C, Halperin Y, Shamir R (2008) Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res* **18:** 1180–1189

Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR, Johnson JM, Cummins JM, Raymond CK, Dai H, Chau N, Cleary M, Jackson AL, Carleton M, Lim L (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27:** 2240–2252

Liu CG, Calin GA, Meloon B, Gamliel N, Sevignani C, Ferracin M, Dumitru CD, Shimizu M, Zupo S, Dono M, Alder H, Bullrich F, Negrini M, Croce CM (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci USA* **101:** 9740–9744

Lodish HF, Zhou B, Liu G, Chen CZ (2008) Micromanagement of the immune system by microRNAs. *Nature Rev* **8:** 120–130

Lu Y, Thomson JM, Wang HY, Hammond SM, Hogan BL (2007) Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev Biol* **310:** 442–453

Milyavsky M, Shats I, Erez N, Tang X, Senderovich S, Meerson A, Tabach Y, Goldfinger N, Ginsberg D, Harris CC, Rotter V (2003) Prolonged culture of telomerase-immortalized human fibroblasts leads to a premalignant phenotype. *Cancer Res* **63:** 7147–7157

Milyavsky M, Tabach Y, Shats I, Erez N, Cohen Y, Tang X, Kalis M, Kogan I, Buganim Y, Goldfinger N, Ginsberg D, Harris CC, Domany E, Rotter V (2005) Transcriptional programs following genetic alterations in p53, INK4A, and H-Ras genes along defined stages of malignant transformation. *Cancer Res* **65:** 4530–4543

Moskovits N, Kalinkovich A, Bar J, Lapidot T, Oren M (2006) p53 attenuates cancer cell migration and invasion through repression of SDF-1/CXCL12 expression in stromal fibroblasts. *Cancer Res* **66:** 10671–10676

Naume B, Zhao X, Synnestvedt M, Borgen E, Russnes EG, Lingjærde OC, Strømberg M, Wiedswang G, Kvalheim G, Kåresen R, Nesland JM, Børresen-Dale AL, Sørlie T (2007) Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol Oncol* **1:** 160–171

O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435:** 839–843

Ookawa K, Tsuchida S, Kohno T, Yokota J (2001) Alterations in expression of E2F-1 and E2F-responsive genes by RB, p53 and p21(Sdi1/WAF1/Cip1) expression. *FEBS Lett* **500:** 25–30

Oren M (2003) Decision making by p53: life, death and cancer. *Cell Death Differ* **10:** 431–442

Ossovskaya VS, Mazo IA, Chernov MV, Chernova OB, Strezoska Z, Kondratov R, Stark GR, Chumakov PM, Gudkov AV (1996) Use of genetic suppressor elements to dissect distinct biological effects of separate p53 domains. *Proc Natl Acad Sci USA* **93:** 10309–10314

Ou YH, Chung PH, Hsu FF, Sun TP, Chang WY, Shieh SY (2007) The candidate tumor suppressor BTG3 is a transcriptional target of p53 that inhibits E2F1. *EMBO J* **26:** 3968–3980

Petrocca F, Visone R, Onelli MR, Shah MH, Nicoloso MS, de Martino I, Iliopoulos D, Pilozzi E, Liu CG, Negrini M, Cavazzini L, Volinia S, Alder H, Ruco LP, Baldassarre G, Croce CM, Vecchione A (2008) E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer. *Cancer Cell* **13:** 272–286

Pilpel Y, Sudarsanam P, Church GM (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* **29:** 153–159

Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N, Bentwich Z, Oren M (2007) Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell* **26:** 731–743

Ren B, Yu G, Tseng GC, Cieply K, Gavel T, Nelson J, Michalopoulos G, Yu YP, Luo JH (2006) MCM7 amplification and overexpression are associated with prostate cancer progression. *Oncogene* **25:** 1090–1098

Ryan KM, Phillips AC, Vousden KH (2001) Regulation and function of the p53 tumor suppressor protein. *Curr Opin Cell Biol* **13:** 332–337

Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, Yuen ST, Chan TL, Kwong DL, Au GK, Liu CG, Calin GA, Croce CM, Harris CC (2008) MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* **299:** 425–436

Shalgi R, Lieber D, Oren M, Pilpel Y (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol* **3:** e131

Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, Borresen-Dale AL (2006) Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics* **7:** 127

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

Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, Ferbeyre G, Chartrand P (2007) An E2F/miR-20a autoregulatory feedback loop. *J Biol Chem* **282:** 2135–2143

Tabach Y, Milyavsky M, Shats I, Brosh R, Zuk O, Yitzhaky A, Mantovani R, Domany E, Rotter V, Pilpel Y (2005) The promoters of human cell cycle genes integrate signals from two tumor suppressive pathways during cellular transformation. *Mol Syst Biol* **1:** 2005.0022

Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. *J Mol Biol* **339:** 327–335

Tarasov V, Jung P, Verdoodt B, Lodygin D, Epanchintsev A, Menssen A, Meister G, Hermeking H (2007) Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G(1)-arrest. *Cell Cycle* **6:** 1586–1593

Tazawa H, Tsuchiya N, Izumiya M, Nakagama H (2007) Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci USA* **104:** 15472–15477

Thai TH, Calado DP, Casola S, Ansel KM, Xiao C, Xue Y, Murphy A, Frendewey D, Valenzuela D, Kutok JL, Schmidt-Supprian M, Rajewsky N, Yancopoulos G, Rao A, Rajewsky K (2007) Regulation of the germinal center response by microRNA-155. *Science* **316:** 604–608

Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu EA (2004) *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303:** 844–848

Vigo E, Muller H, Prosperini E, Hateboer G, Cartwright P, Moroni MC, Helin K (1999) CDC25A phosphatase is a target of E2F and is required for efficient E2F-induced S phase. *Mol Cell Biol* **19:** 6379–6395

Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* **103:** 2257–2261

Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, Zlotorynski E, Yabuta N, De Vita G, Nojima H, Looijenga LH, Agami R (2006) A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124:** 1169–1181

Wang Q, Li YC, Wang J, Kong J, Qi Y, Quigg RJ, Li X (2008) miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130. *Proc Natl Acad Sci USA* **105:** 2889–2894

Whitfield ML, George LK, Grant GD, Perou CM (2006) Common markers of proliferation. *Nat Rev Cancer* **6:** 99–106

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13:** 1977–2000

Woods K, Thomson JM, Hammond SM (2007) Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *J Biol Chem* **282:** 2130–2134

Wu H, Neilson JR, Kumar P, Manocha M, Shankar P, Sharp PA, Manjunath N (2007) miRNA profiling of naive, effector and memory CD8 T Cells. *PLoS ONE* **2:** e1020

Xi Y, Shalgi R, Fodstad O, Pilpel Y, Ju J (2006) Differentially regulated micro-RNAs and actively translated messenger RNA transcripts by tumor suppressor p53 in colon cancer. *Clin Cancer Res* **12:** 2014–2024

Xue W, Zender L, Miething C, Dickins RA, Hernando E, Krizhanovsky V, Cordon-Cardo C, Lowe SW (2007) Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445:** 656–660

Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, Harris CC (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9:** 189–198

Table of Content

Figure S1

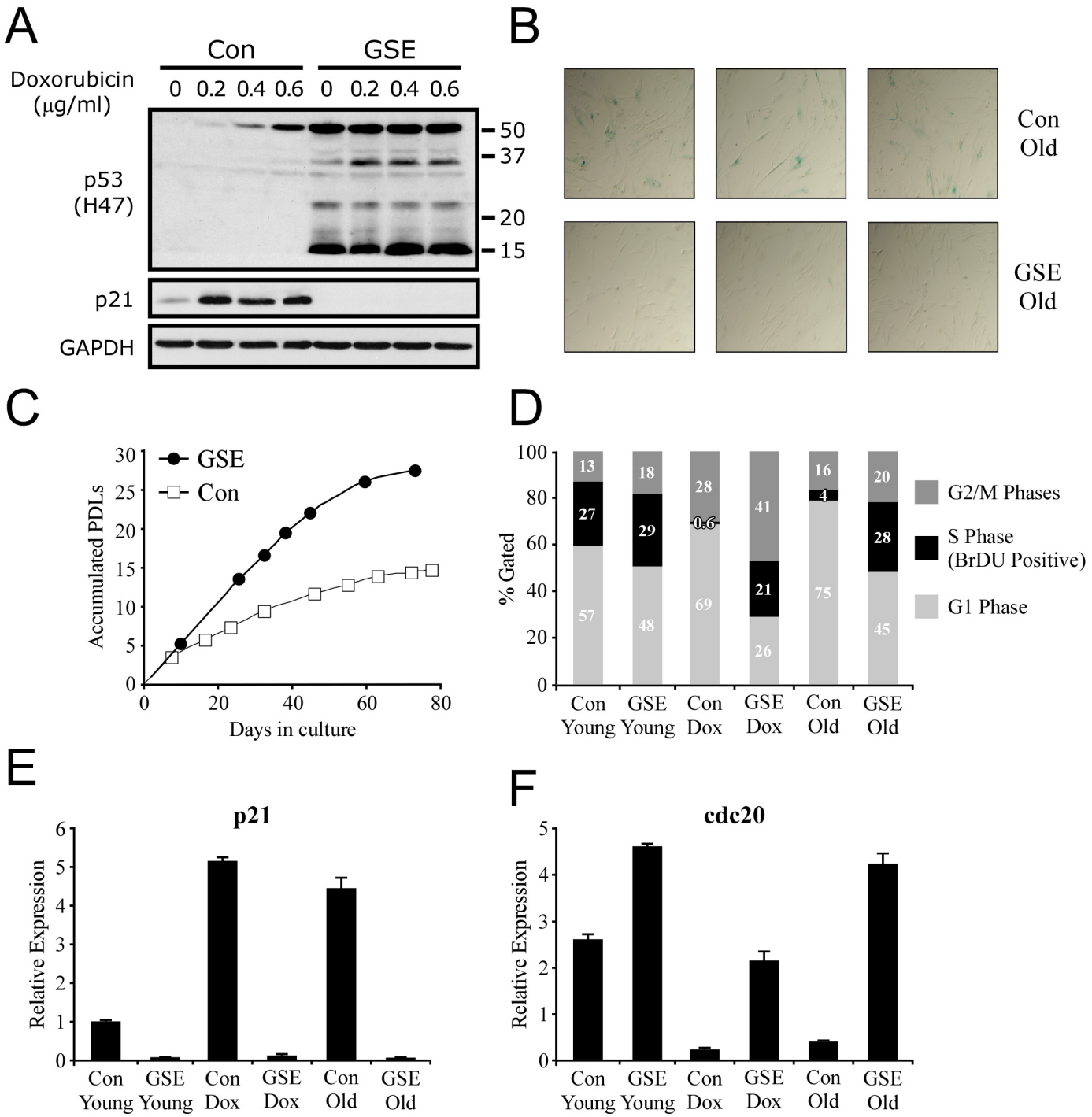

A
Con
GSE
Doxorubicin (μg/ml)
0 0.2 0.4 0.6 0 0.2 0.4 0.6
p53 (H47)
p21
GAPDH
50
37
20
15
B
Con Old
GSE Old
C
GSE
Con
Accumulated PDLs
Days in culture
D
% Gated
G2/M Phases
S Phase (BrDU Positive)
G1 Phase
Con Young
GSE Young
Con Dox
GSE Dox
Con Old
GSE Old
E
p21
Relative Expression
F
cdc20
Relative Expression

Figure S1. **Establishment of the WI-38 system.**
WI-38 primary human fibroblasts were infected with a retrovirus encoding for the p53-inactivating peptide, GSE56. These cells (GSE) and their active p53 counterparts (Con) were treated with the DNA damaging agent doxorubicin as well as grown until the onset of replicative senescence. A. Western blot depicting p53 and p21 following doxorubicin treatment. p53 was stabilized and activated its target gene p21 in the NEO cells upon treatment. In the GSE cells, p53 was stabilized in its inactive form already at basal levels by GSE56, a 15 kDa peptide detected by the H-47 anti-p53 polyclonal antibody. The p21 protein was not detectable in the GSE cells at basal levels, nor was it induced upon DNA damage, indicating of a complete inactivation of p53 transactivation activity. B and C. Introduction of the GSE56 peptide at passage 20 resulted in accelerated growth rate (C) and delay in replicative senescence, as indicated by the reduction of senescence-associated -galactosidase (SA- -Gal) staining (B). D. Cell cycle analysis demonstrates that both DNA damage and replicative senescence resulted in a sharp p53-dependent cell cycle arrest. E. QRT-PCR analyses of p21 mRNA levels demonstrated that p53 transactivation activity was significantly induced upon both DNA damage and replicative senescence, and was completely abolished by the introduction of GSE56. F. QRT-PCR analyses of cdc20, a p53-repressed gene that participates in cell-cycle progression. To conclude, we have generated an isogenic pair of primary human cell cultures that display p53-dependent cell cycle arrest, gene activation and repression, and senescence phenotypes upon application of p53-activating stress. Data are represented as mean ± SD.

Figure S2

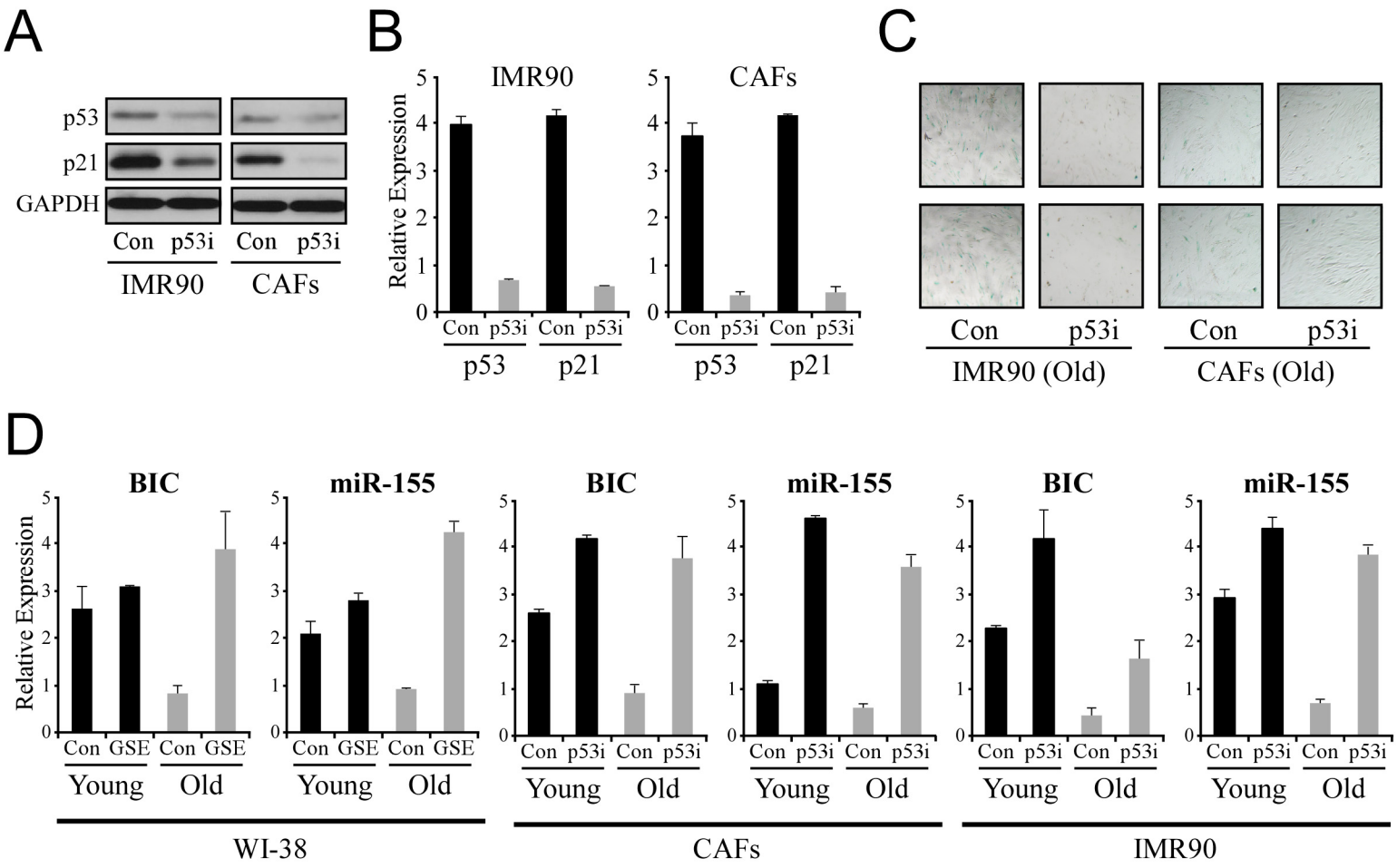
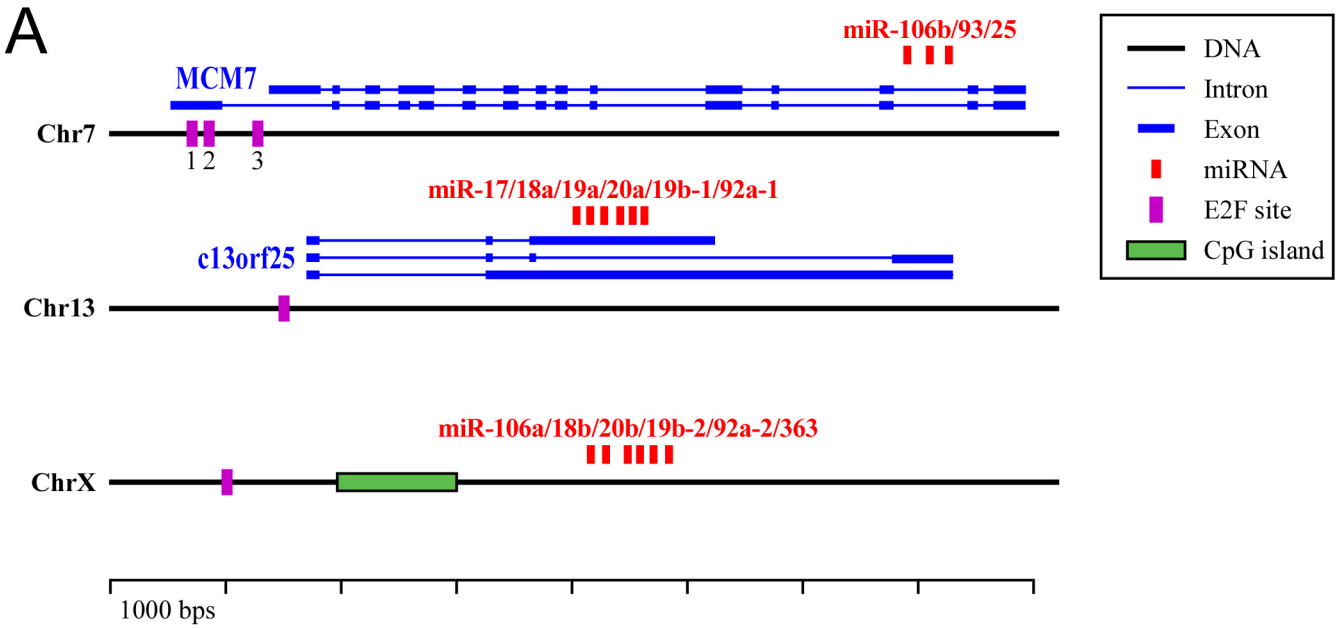
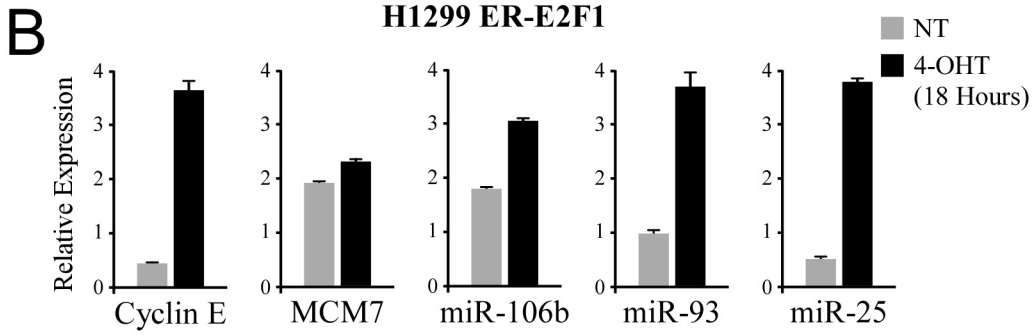
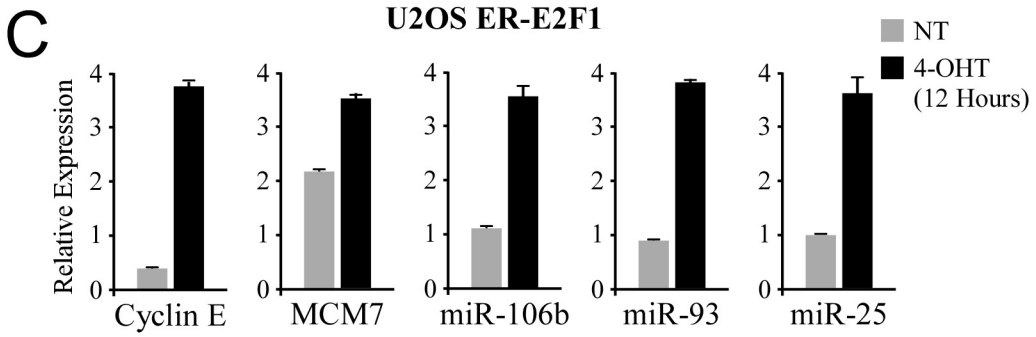


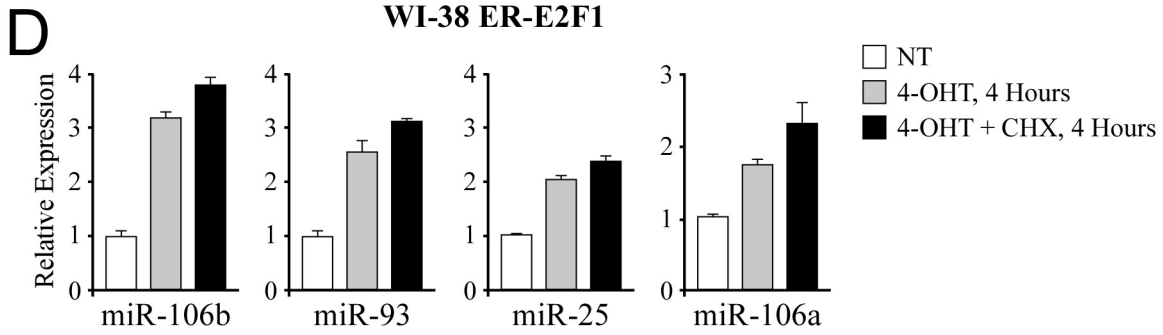
Figure S2. **Establishment of the IMR90 and CAFs systems.** Lung primary human fibroblasts IMR90 and prostate-cancer associated fibroblasts (CAFs) PFCA179 were infected with a retrovirus encoding for either a small hairpin RNA targeting p53 (p53i) or a control RNAi (Con), and grown until the onset of replicative senescence. A. A Western blot depicting p53 and p21 downregulation upon the stable expression of the p53 small hairpin RNA. B. QRT-PCR analyses of p53 and p21 mRNA levels. C. SA- -Gal staining for late passage IMR90 and CAFs. D. QRT-PCR analysis for the non-coding RNA BIC and its resident miRNA miR-155 in WI-38, CAFs and IMR90 cells. Samples were collected from early passage cultures (Young) and from late passage cultures (Old). Data are represented as mean ± SD.

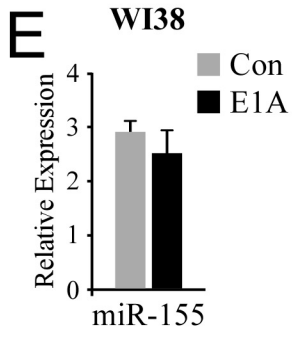
Figure S3


A
miR-106b/93/25
MCM7
Chr7
1 2 3
miR-17/18a/19a/20a/19b-1/92a-1
c13orf25
Chr13
miR-106a/18b/20b/19b-2/92a-2/363
ChrX
1000 bps
DNA
Intron
Exon
miRNA
E2F site
CpG island


B
H1299 ER-E2F1
NT
4-OHT
(18 Hours)
Relative Expression
Cyclin E
MCM7
miR-106b
miR-93
miR-25


C
U2OS ER-E2F1
NT
4-OHT
(12 Hours)
Relative Expression
Cyclin E
MCM7
miR-106b
miR-93
miR-25


D
WI-38 ER-E2F1
NT
4-OHT, 4 Hours
4-OHT + CHX, 4 Hours
Relative Expression
miR-106b
miR-93
miR-25
miR-106a


E
WI38
Con
E1A
Relative Expression
miR-155

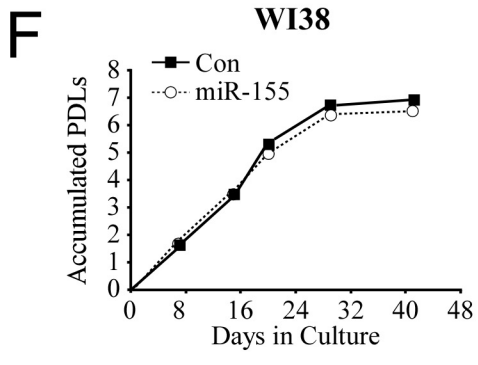

F
WI38
Con
miR-155
Accumulated PDLs
Days in Culture

Figure S3. The cell-cycle associated polycistronic miRNAs are directly induced by E2F.

A. Schematic representation of the genomic organization of the three paralogous polycistrons. Conserved E2F sites (according to the tfbsConsSites track from the UCSC database) are located upstream of each of the miRNA clusters. For the miR-106b-25 cluster, three E2F sites are found within 1kb upstream of MCM7 TSS. For miR-17-92 cluster, an E2F site is located 83 bps upstream of c13orf25 TSS. For miR-106a-92 cluster, which has no known host transcript, an E2F site is located upstream of a CpG island, indicative of a potential TSS in that region. ChIP analysis indicates binding of E2F1 to each of the E2F sites presented here (Fig. 4A). B and C. E2F induces the levels of miR-106b/93/25 polycistron. H1299 lung carcinoma cell line (A) and U2OS osteosarcoma cell line (B) were treated with 4-OHT for the indicated time periods. QRT-PCR analyses demonstrated upregulation of the known E2F1 target, Cyclin E; as well as of MCM7 and its resident miRNAs. D. ER-E2F1 expressing WI-38 cells were treated with 4-OHT (300nM) in combination with 80 g/ml of cycloheximide (CHX) for 4 hours. The levels of the polycistronic miRNAs were measured by QRT-PCR. Note that cycloheximide did not attenuate the induction of the miRs by 4-OHT, indicating that the transactivation of the miRs by E2F1 does not require the synthesis of a protein mediator. E. WI-38 cells were infected with the oncoprotein E1A or a control vector (Con) and QRT-PCR analysis was performed to measure the level of miR-155. F. Growth curves for control (empty vector) and miR-155 overexpressing cells. Both cell types displayed comparable growth rate and entered replicative senescence simultaneously. PDLs = Population Doublings. Data are represented as mean ± SD.

Figure S4

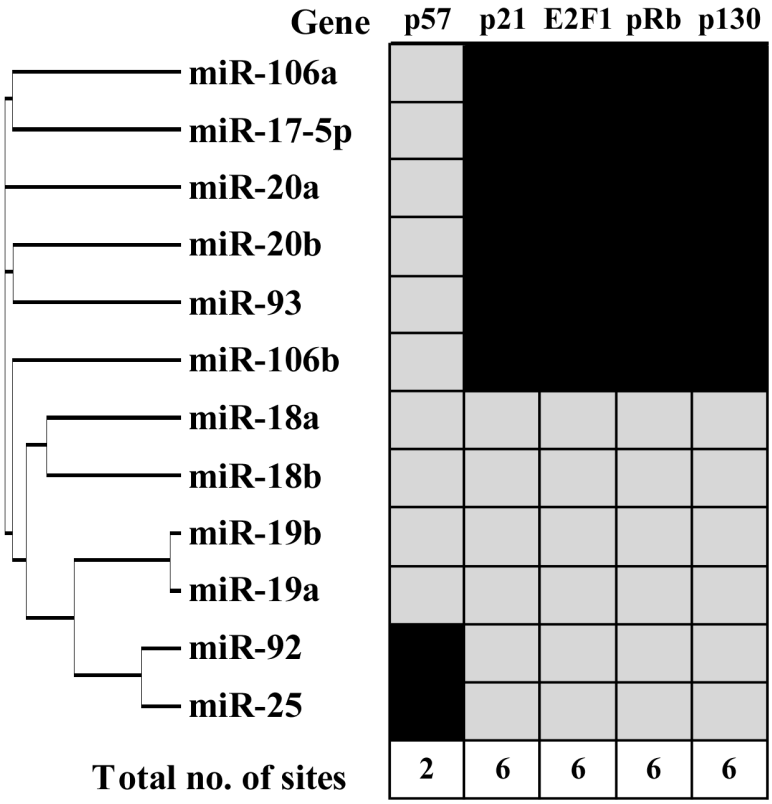
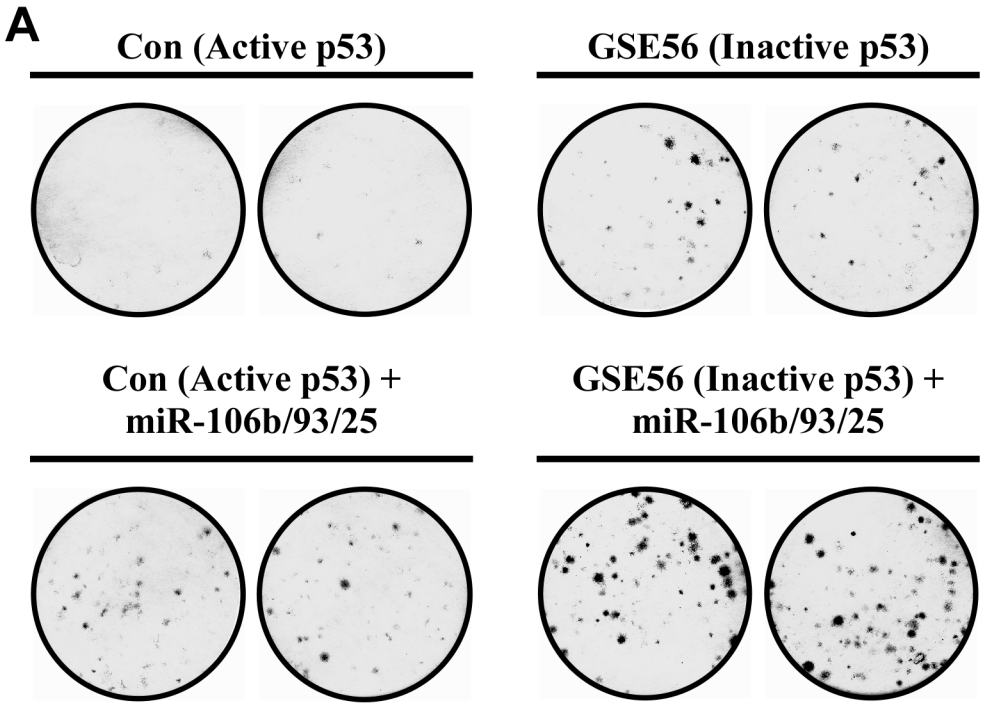


Figure S4. **miRNAs from the three paralogous polycistrons and their cell-cycle associated targets.**

Targeting of cell-cycle associated genes by miRNAs that belong to the miR-17-93, miR-106a-93 and miR-106b-25 polycistrons as predicted by PicTar. Black areas indicate predicted targeting.

Figure S5


A
Con (Active p53)
GSE56 (Inactive p53)
Con (Active p53) +
miR-106b/93/25
GSE56 (Inactive p53) +
miR-106b/93/25

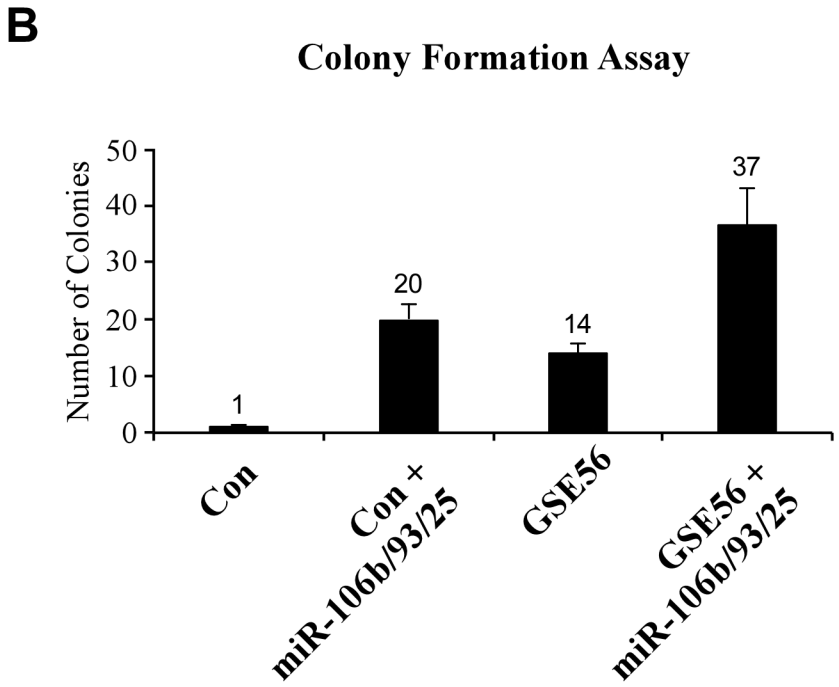

B
Colony Formation Assay
Number of Colonies
50
40
30
20
10
0
1
20
14
37
Con
Con +
miR-106b/93/25
GSE56
GSE56 +
miR-106b/93/25

Figure S5. **Effect of miR-106b/93/25 on colony formation efficiency in WI-38 cells in the presence and absence of GSE56.** WI-38 primary human fibroblasts were infected with a retrovirus encoding for the p53-inactivating peptide, GSE56. These cells and their active-p53 counterparts (Con) were infected with a retrovirus encoding for either the genomic region that contains miR-106b/93/25 or an empty vector control. After selection with the appropriate selection drug cells were plated (in duplicates) at low density and incubated for 10 days. A. Colonies were stained with crystal violet and plates were scanned. B. Colonies were counted. The average number of colonies (and standard deviation) for each condition are plotted. Note that the effect of miR-106b/93/25 overexpression in the control cells (that express an active p53) is more pronounced than in the GSE56-expressing cells.

Figure S6

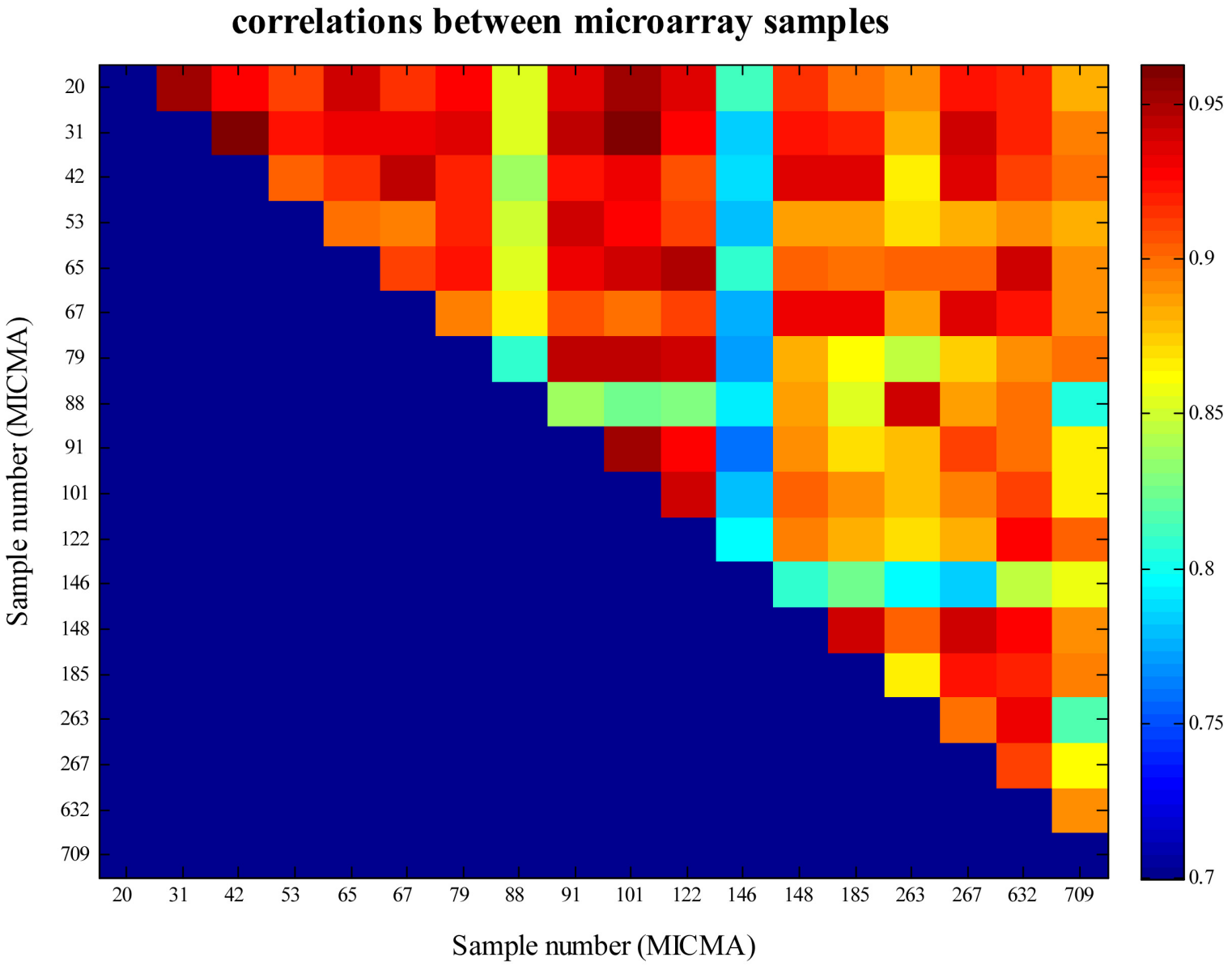


Figure S6. **Correlation coefficient between pairs of miRNA microarrays from the breast cancer study.**

Two samples, MICMA88 and MICMA146, were consistently more different from all other samples. Therefore, they were discarded from subsequent clustering analysis. These samples represent two out of 12 tumor samples with mutant p53.

Table S1. Detailed description of TP53 gene mutations and expression subtypes of the 16 breast cancer samples

Samples series: MicMa; Material: DNA, tumor; Method: Sequencing, TTGE; Region analysed: Exon 2-11; Reference sequence: NM_095720.

Sample ID	Expression Subgroup	TP53 status	Mutated Exon	Codon	Codon Change	Substitution	Mutation Type (predicted)	Codon 72 Polymorphism	Other variants, Polymorphism
31	Basal like	MUT	7	234	TAC>TGC	Tyr>Cys	Missense	G/G Arg/Arg	
42	Basal like	MUT	7	248	CGG>CAG	Arg>Gln	Missense	G/C Arg/Pro	c213, CGA>CGG, Arg>Arg,het
53	ERBB2	MUT	5	141	TGC>TAC	Cys>Tyr	Missense	G/G Arg/Arg	
67	Basal like	MUT	8	274	GTT>TTT	Val>Phe	Missense	G/G Arg/Arg	
79	ERBB2	MUT	6	209	Deletion of 2 bps		Frameshift	G/G Arg/Arg	
91	Luminal B	MUT	9		IVS9+1 G>A		Splice	G/C Arg/Pro	IVS4-29C>A,het
148	Luminal B	MUT	8	274	GTT>GAT	Val>Asp	Missense	G/G Arg/Arg	
185	Basal like	MUT	5	173	GTG>ATG	Val>Met	Missense	G/G Arg/Arg	
267	Basal like	MUT	6	213	CGA>TGA	Arg>Stop	Nonsense	G/G Arg/Arg	
709	Basal like	MUT	6	195/196	Deletion of 1 bp		Frameshift	G/C Arg/Pro	
20	Normal like	WT						G/G Arg/Arg	
65	Luminal A	WT						G/C Arg/Pro	
101	Luminal A	WT						G/G Arg/Arg	
122	Luminal A	WT						G/C Arg/Pro	
263	Luminal A	WT						G/G Arg/Arg	
632	Luminal A	WT						G/C Arg/Pro	

Table S2. miRNAs that were co-clustered with mRNAs enriched for distinct functional annotations

Cell-Cycle	Immune Response
miR-106a	miR-142-3p
miR-106b	miR-142-5p
miR-135b	miR-146b
miR-146a	miR-150
miR-15b	miR-155
miR-17-3p	miR-7
miR-17-5p	
miR-18a	
miR-18b	
miR-19a	
miR-19b	
miR-20a	
miR-20b	
miR-223	
miR-25	
miR-32	
miR-362	
miR-363	
miR-454-3p	
miR-500	
miR-502	
miR-532	
miR-545	
miR-576	
miR-579	
miR-590	
miR-660	
miR-9	
miR-9*	
miR-92	
miR-93	

Table S3. Functional annotation enrichment analysis for clustered genes derived from breast cancer samples.

A

Functional Annotation Term	No. of Cluster Genes	Enrichment P-Value
Mitotic cell cycle	39	8.53E-20
Cell cycle	70	1.11E-18
M phase	34	2.62E-16
Mitosis	28	3.13E-14
M phase of mitotic cell cycle	28	4.39E-14
Cell division	27	1.37E-11
Chromosome segregation	14	2.63E-11
Regulation of progression through cell cycle	40	4.86E-09
Regulation of cell cycle	40	5.16E-09
Cell cycle checkpoint	12	2.20E-08
Cell proliferation	41	8.27E-08
Cytoskeleton organization and biogenesis	34	1.09E-07
Organelle organization and biogenesis	53	3.77E-07
DNA metabolism	42	8.93E-06
Phosphoinositide-mediated signaling	12	1.76E-05
Mitotic sister chromatid segregation	7	2.24E-05
Sister chromatid segregation	7	3.01E-05
Microtubule-based process	17	4.50E-05
Interphase of mitotic cell cycle	10	5.62E-05
Interphase	10	7.32E-05
Response to DNA damage stimulus	21	8.10E-05
DNA-dependent DNA replication	12	8.23E-05
Ectoderm development	12	9.08E-05

B

Functional Annotation Term	No. of Cluster Genes	Enrichment P-Value
Immune response	92	1.24E-48
Response to biotic stimulus	97	2.12E-48
Defense response	94	5.36E-47
Response to pest, pathogen or parasite	56	2.52E-32
Response to other organism	56	5.97E-31
Response to stimulus	106	1.08E-28
Organismal physiological process	99	9.36E-27
Response to stress	60	4.50E-21
Response to virus	17	8.29E-15
Response to external stimulus	35	1.23E-14
Response to wounding	30	9.15E-14
Humoral immune response	20	7.40E-13
Inflammatory response	20	6.76E-11
Chemotaxis	16	1.20E-10
Taxis	16	1.20E-10
Locomotory behavior	16	2.10E-10
Regulation of apoptosis	24	4.29E-10
Regulation of programmed cell death	24	4.80E-10
Humoral defense mechanism	15	5.97E-10
Antimicrobial humoral response	13	2.77E-09
Cell death	29	4.28E-09
Death	29	4.96E-09
Apoptosis	28	8.01E-09
Programmed cell death	28	8.63E-09
Behavior	17	9.69E-09
Antimicrobial humoral response	12	2.03E-08
Calcium ion homeostasis	11	6.99E-08
Response to chemical stimulus	22	7.35E-08
Response to abiotic stimulus	23	1.84E-07
Lymphocyte activation	11	4.28E-07
Cell communication	79	5.67E-07
Signal transduction	74	8.96E-07
Cellular defense response	11	1.33E-06
Di-, tri-valent inorganic cation homeostasis	11	2.04E-06
Immune cell activation	11	2.22E-06
Cell activation	11	2.41E-06
Metal ion homeostasis	11	3.88E-06
Induction of programmed cell death	12	4.01E-06

Table S4. Conserved E2F sites predictions upstream of the three miRNA polycistrons

Predicted E2F sites were searched for in the genomic region spanning up to 10,000 bps upstream of the first miRNA in each polycistron (i.e. miR-17-5p, miR-106b and miR-106a), using the *tfbsConsSites database* (UCSC, human genome version hg17).

miRNA Polycistron	Chr	Site name	Position-Specific Scoring Matrix (PSSM)	Distance Upstream of first miRNA in the polycistron	Distance upstream from Primary Transcript/Host TSS	Site Sequence
hsa-mir-17-92	13	miR-17-92 site 1	V$E2F_Q3	2868	83	CCTTCGCGC
hsa-mir-106b-25	7	miR-106b-25 site 1	V$E2F_01	7513	814	ACCGCGGGAAACCCGG
hsa-mir-106b-25	7	miR-106b-25 site 2	V$E2F_01	7336	637	GACGTTTCGCGCCAAT
hsa-mir-106b-25	7	miR-106b-25 site 3	V$E2F4DP2_01	6821	129	GTTCCCGCG
hsa-mir-106a-92	X	miR-106a-92 site 1	V$E2F_Q3	3798	no known primary transcript	CTTCGCGCC

Table S5. Mutual targets of E2F transactivation and miR-106b/93/25 silencing

The list of E2F targets was compiled from a combination of seven high-throughput studies designed to identify E2F target genes (Ishida et al., 2001; Ma et al., 2002; Muller et al., 2001; Polager et al., 2002; Ren et al., 2002; Stanelle et al., 2002; Weinmann et al., 2001). Prediction of miRNA sites was performed using PicTar (Krek et al., 2005).

RefSeq	Symbol	Name	PicTar predicted sites		
NM_000076	CDKN1C	cyclin-dependent kinase inhibitor 1C	miR-25		
NM_000321	RB1	retinoblastoma 1	miR-93	miR-106b	
NM_001124	ADM	adrenomedullin	miR-25		
NM_001386	DPYSL2	dihydropyrimidinase-like 2	miR-93	miR-106b	
NM_001396	DYRK1A	dual-specificity tyrosine-(Y)-phosphorylation	miR-93	miR-106b	
NM_001430	EPAS1	endothelial PAS domain protein 1	miR-93	miR-106b	
NM_001450	FHL2	four and a half LIM domains 2 isoform 1	miR-25		
NM_001753	CAV1	caveolin 1	miR-106b		
NM_001769	CD9	CD9 antigen	miR-25		
NM_001949	E2F3	E2F transcription factor 3	miR-93	miR-106b	miR-25
NM_002024	FMR1	fragile X mental retardation 1	miR-25		
NM_002499	NEO1	neogenin homolog 1	miR-93		
NM_003016	SFRS2	splicing factor, arginine/serine-rich 2	miR-93	miR-106b	
NM_003139	SRPR	signal recognition particle receptor ('docking	miR-25		
NM_003272	TM7SF1	transmembrane 7 superfamily member 1	miR-106b		
NM_003505	FZD1	frizzled 1	miR-106b		
NM_003954	MAP3K14	mitogen-activated protein kinase kinase kinase	miR-93	miR-106b	
NM_004036	ADCY3	adenylate cyclase 3	miR-25		
NM_004354	CCNG2	cyclin G2	miR-93	miR-106b	

NM_004364	CEBPA	CCAAT/enhancer binding protein alpha	miR-25		
NM_004844	SH3BP5	SH3-domain binding protein 5 (BTK-associated)	miR-93	miR-106b	
NM_005197	CHES1	checkpoint suppressor 1	miR-106b		
NM_005458	GPR51	G protein-coupled receptor 51	miR-106b		
NM_005596	NFIB	nuclear factor I/B	miR-93	miR-25	
NM_005779	LHFPL2	lipoma HMGIC fusion partner-like 2	miR-25		
NM_005923	MAP3K5	mitogen-activated protein kinase kinase kinase	miR-93	miR-106b	
NM_006195	PBX3	pre-B-cell leukemia transcription factor 3	miR-106b		
NM_006352	ZNF238	zinc finger protein 238 isoform 2	miR-93	miR-106b	
NM_006482	DYRK2	dual-specificity tyrosine-(Y)-phosphorylation	miR-93	miR-25	
NM_006751	SSFA2	sperm specific antigen 2	miR-93	miR-106b	miR-25
NM_006806	BTG3	B-cell translocation gene 3	miR-93	miR-106b	
NM_012334	MYO10	myosin X	miR-106b		
NM_014344	FJX1	four jointed box 1	miR-93	miR-106b	
NM_014679	KIAA0092	translokin	miR-93	miR-106b	
NM_014876	KIAA0063	KIAA0063 gene product	miR-93	miR-106b	miR-25
NM_015008	KIAA0779	KIAA0779 protein	miR-93	miR-106b	
NM_015678	NBEA	neurobeachin	miR-93	miR-106b	
NM_016131	RAB10	ras-related GTP-binding protein RAB10	miR-93	miR-106b	
NM_016357	EPLIN	epithelial protein lost in neoplasm beta	miR-93	miR-106b	
NM_017803	FLJ20399	hypothetical protein FLJ20399	miR-25		
NM_021005	NR2F2	nuclear receptor subfamily 2, group F, member 2	miR-106b		
NM_053056	CCND1	cyclin D1	miR-93	miR-106b	
NM_057749	CCNE2	cyclin E2 isoform 1	miR-106b	miR-25	
NM_145725	TRAF3	TNF receptor-associated factor 3 isoform 1	miR-25		
NM_153719	NUP62	nucleoporin 62kDa	miR-93		

NM_173075	APBB2	amyloid beta A4 precursor protein-binding,	miR-93	miR-106b	
NM_173173	NR4A2	nuclear receptor subfamily 4, group A, member 2	miR-93	miR-106b	
NM_173200	NR4A3	nuclear receptor subfamily 4, group A, member 3	miR-93	miR-106b	miR-25
NM_174886	TGIF	TG-interacting factor isoform d	miR-25		
NM_181659	NCOA3	nuclear receptor coactivator 3 isoform a	miR-93	miR-106b	
NM_182744	NBL1	neuroblastoma, suppression of tumorigenicity 1	miR-93	miR-106b	
NM_183384	RNF13	ring finger protein 13 isoform 3	miR-106b		
NM_197968	ZNF198	zinc finger protein 198	miR-93		
NM_198966	PTHLH	parathyroid hormone-like hormone isoform 1	miR-93	miR-106b	
NM_199334	THRA	thyroid hormone receptor, alpha isoform 1	miR-93	miR-106b	
NM_024322	MGC11266	hypothetical protein MGC11266	miR-93	miR-106b	

Ishida, S., Huang, E., Zuzan, H., Spang, R., Leone, G., West, M., and Nevins, J.R. (2001). Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. Mol Cell Biol *21*, 4684-4699.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M.*, et al.* (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.

Ma, Y., Croxton, R., Moorer, R.L., Jr., and Cress, W.D. (2002). Identification of novel E2F1-regulated genes by microarray. Arch Biochem Biophys *399*, 212-224.

Muller, H., Bracken, A.P., Vernell, R., Moroni, M.C., Christians, F., Grassilli, E., Prosperini, E., Vigo, E., Oliner, J.D., and Helin, K. (2001). E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. Genes Dev *15*, 267-285.

Polager, S., Kalma, Y., Berkovich, E., and Ginsberg, D. (2002). E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. Oncogene *21*, 437-446.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev *16*, 245-256.

Stanelle, J., Stiewe, T., Theseling, C.C., Peter, M., and Putzer, B.M. (2002). Gene expression changes in response to E2F1 activation. Nucleic Acids Res *30*, 1859-1867.

Weinmann, A.S., Bartley, S.M., Zhang, T., Zhang, M.Q., and Farnham, P.J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol Cell Biol *21*, 6820-6832.

Table S6A. Primers used for quantification of mRNA levels (using QRT-PCR).

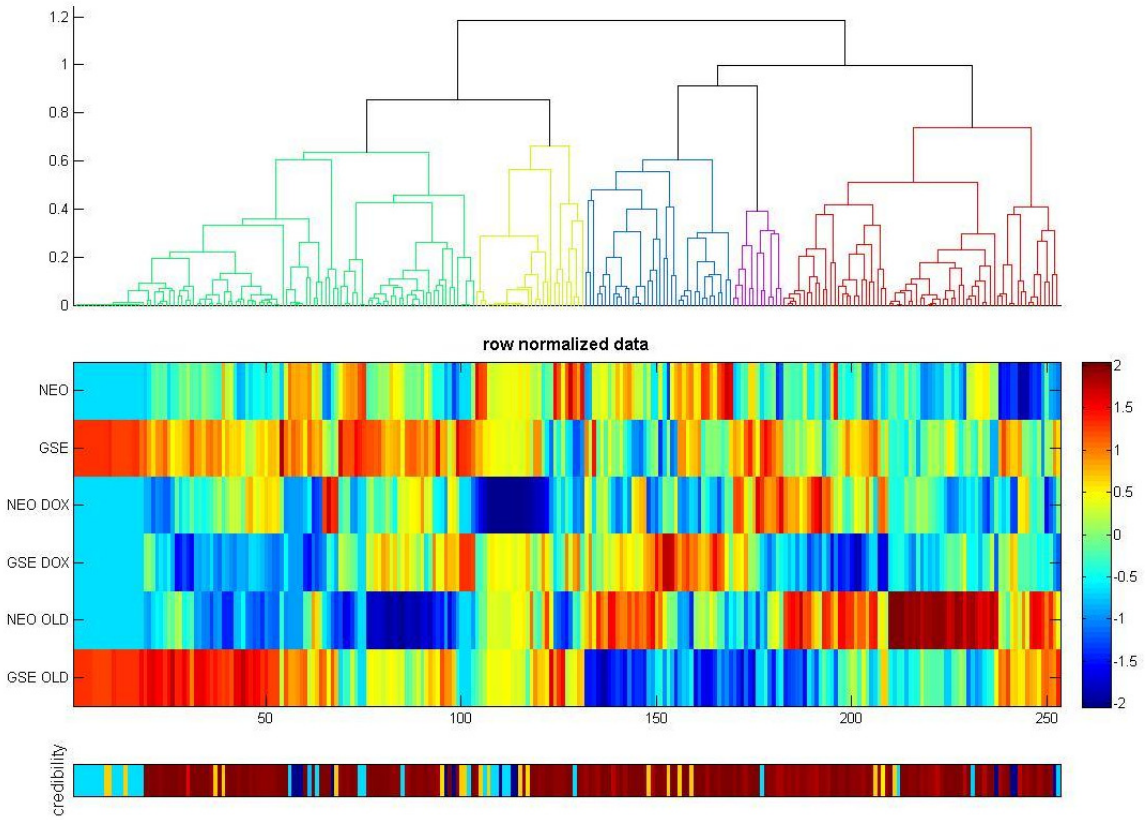
Gene Symbol , RefSeq No.	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
GAPDH, NM_002046	agcctcaagatcatcagcaatg	cacgataccaaagttgtcatggat
MCM7, NM_005916	tctggcacgtctgagaatggt	acggacggtggcaaatatca
C13ORF25	gaagatggtggcggctactc	ggtgcagttaggtccacgtgtat
E2F1, NM_005225	ccatccaggaaaaggtgtgaa	agcgcttggtggtcagattc
CCNE1 (Cyclin E), NM_001238	tttacccaaactcaacgtgcaa	tcggagacctaccacgttatta
CDKN1A (p21), NM_078467	cgcgactgtgatgcgctaatg	ggaacctctcattcaaccgcc
CDKN1C (p57), NM_000076	gaacgccgaggaccagaac	ggcatgtcctgctggaagtc
RBL2 (p130), NM_005611	caagcaacctcagccttcca	ttctctccatctaaagttaccgaaga
CDC20, NM_001255	gagggtggctgggttcctct	cagatgcgaatgtgtcgatca
BIC	tgtgcgagcagagaatctaccttt	tggaggaagaaacaggcttagaa

Table S6B. Primers used for ChIP analysis (using QRT-PCR).

Primer pair name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
miR-106b-25 sites 1-2	gtgattggcttgcggctag	caatcggacaaggcggc
miR-106b-25 site 3	tcttaagggctctgggctcc	ggaatgcccaaaagcgc
miR-17-92	ttttatgctaatgagggagtggg	gctcccgcctcaacgtaa
miR-106a-92	gctgcagctgtaggacacaattaat	gctacatccgctcctcacaaa
β-actin	actggctcgtgtgacaaggc	cactccaaggccgctttaca
β-tubulin (NM_177987)	ggagctgatggagtcagtgatg	cagctctcagcctcctttctg

Supplementary Information

Clustering Analysis - WI-38 Young vs. senescent, p53 inactivation with GSE

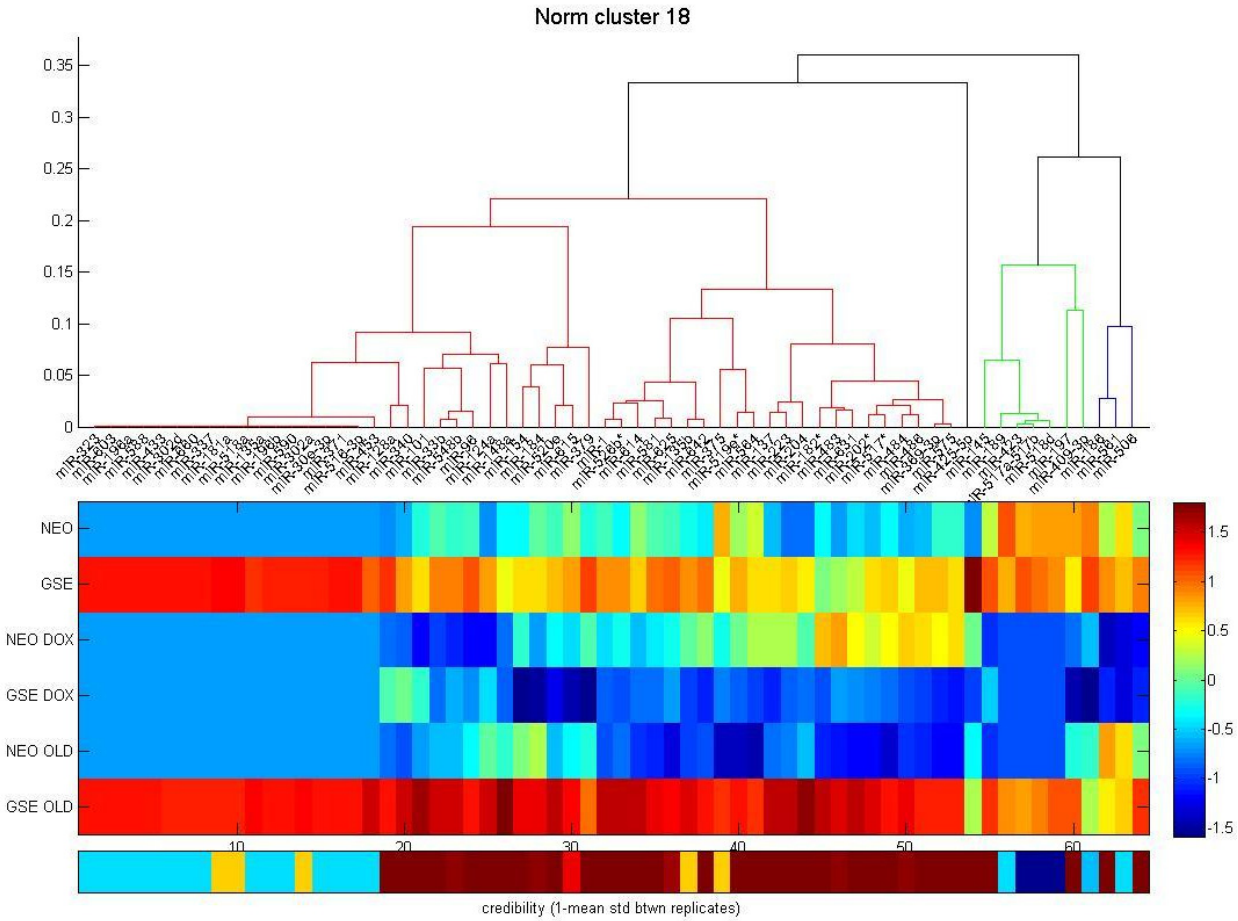


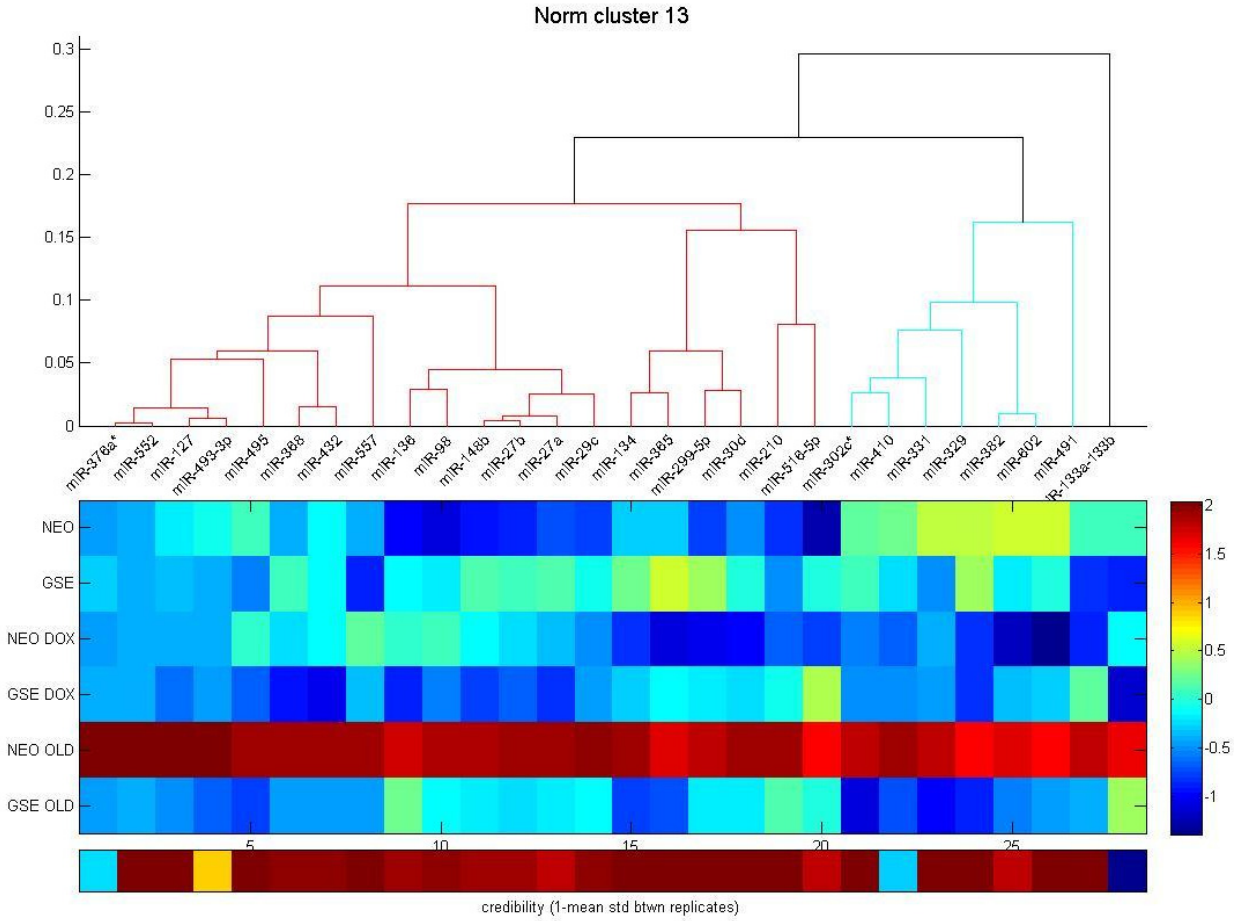
Dendrogram of the entire dataset.

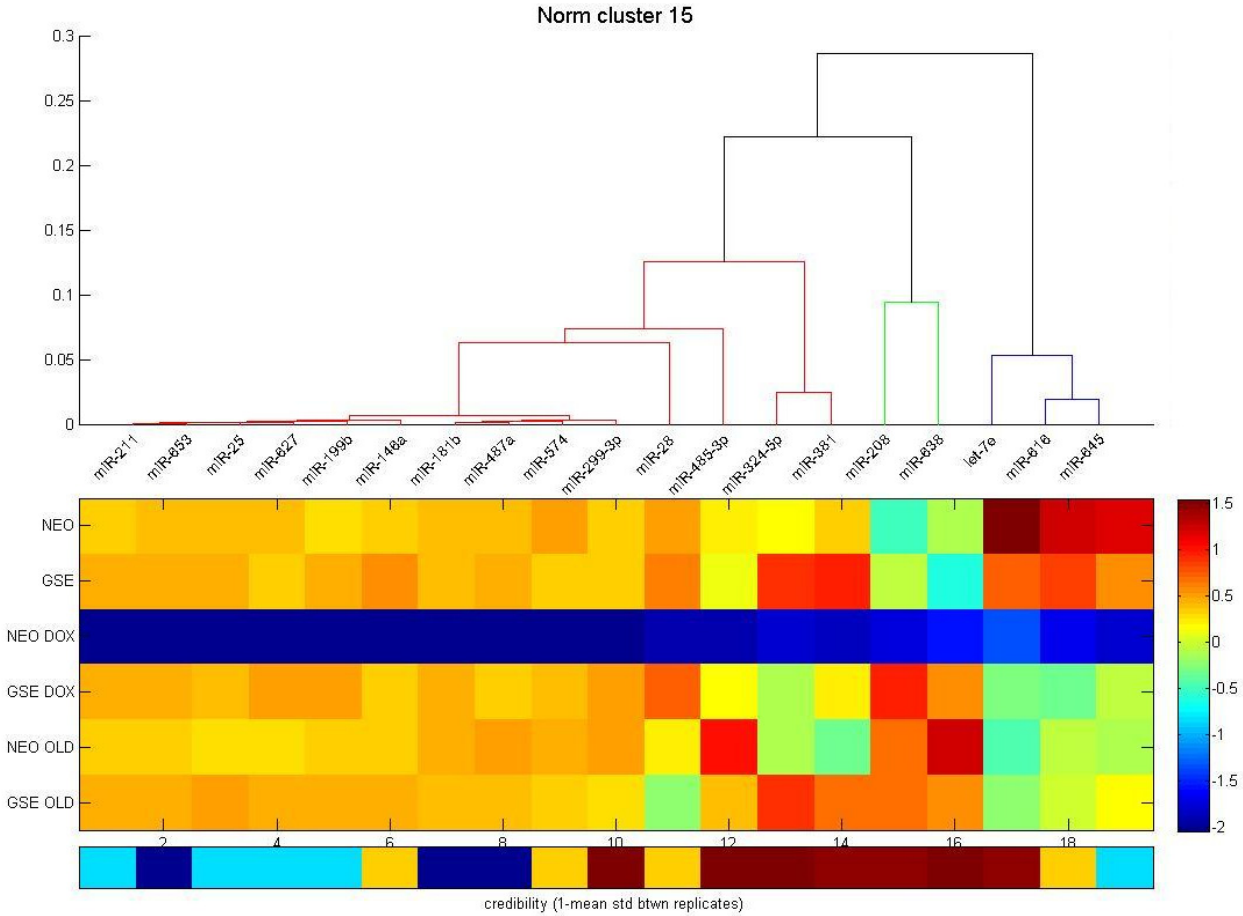
Hy5/Hy3 ratios were log2 transformed and filtered such that miRs which were undetected in 11 or 12 samples were discarded. Duplicates were averaged, such that each miR was represented by six values,
corresponding to the six different samples. For each miR, a credibility value was calculated as one minus the average of the six standard deviations (SD) between the duplicates. A duplicate that had one missing value was set as the detected value and was assigned with high SD. The 5% most non-credible miRs were discarded. Data was clustered using hierarchical clustering (average linkage), with 20 clusters.

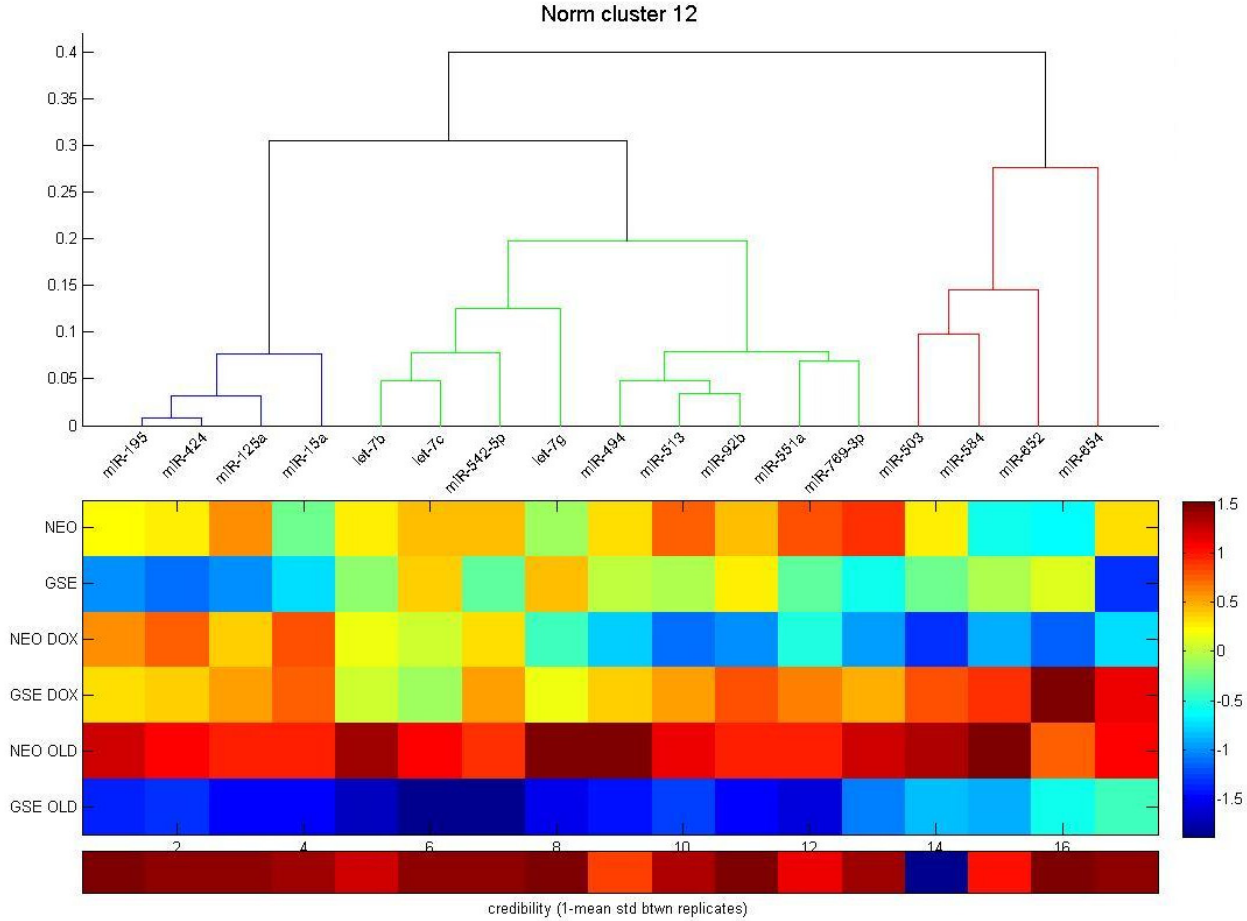
Additional Clusters

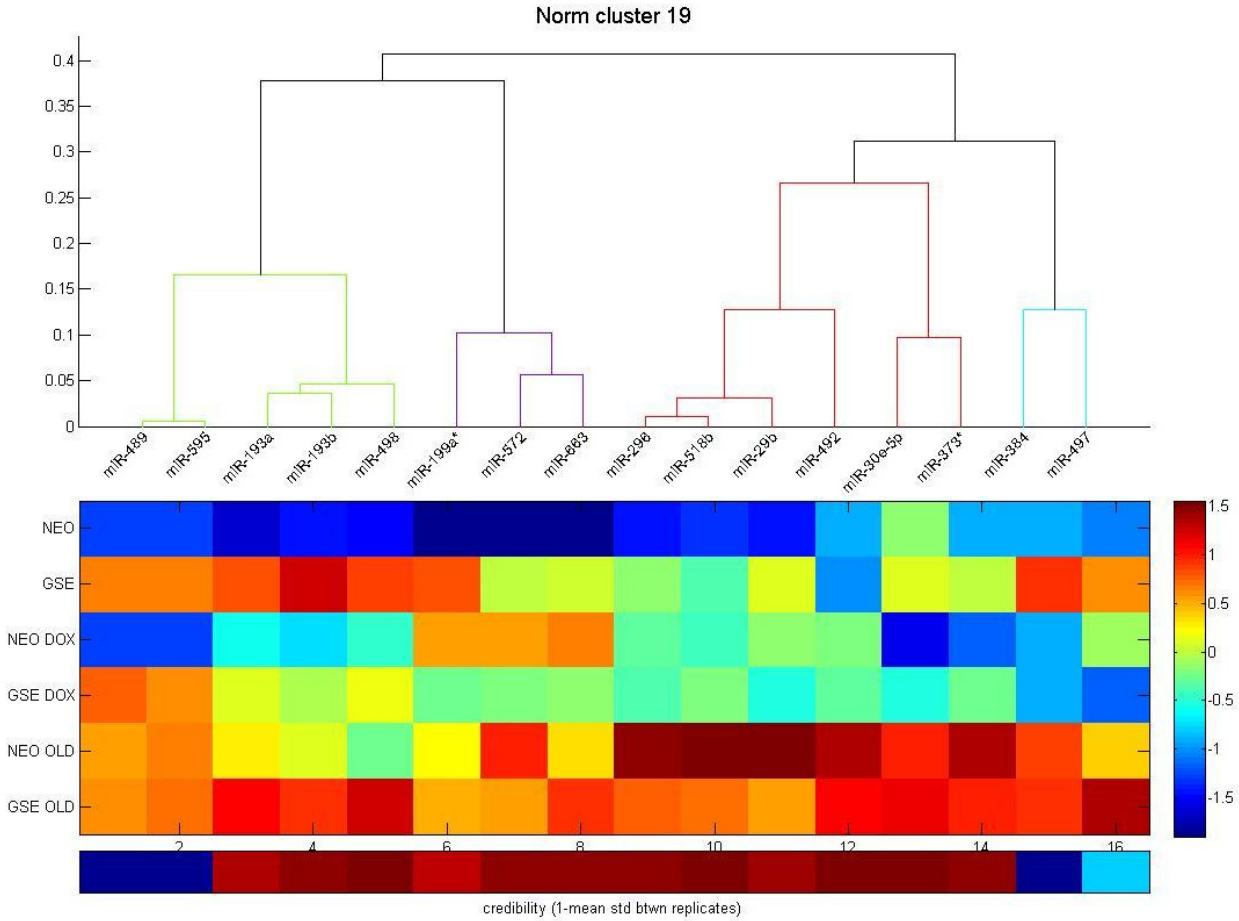
Below we present additional clusters out of the 20, which resulted from the clustering analysis of the data above.
miRNA names, and a dendrogram of their expression similarity, normalized expression data (minus mean, divided by std for each miR) and credibility values are presented for all miRs in each cluster.
We present here clusters with more than 10 miRs in them/

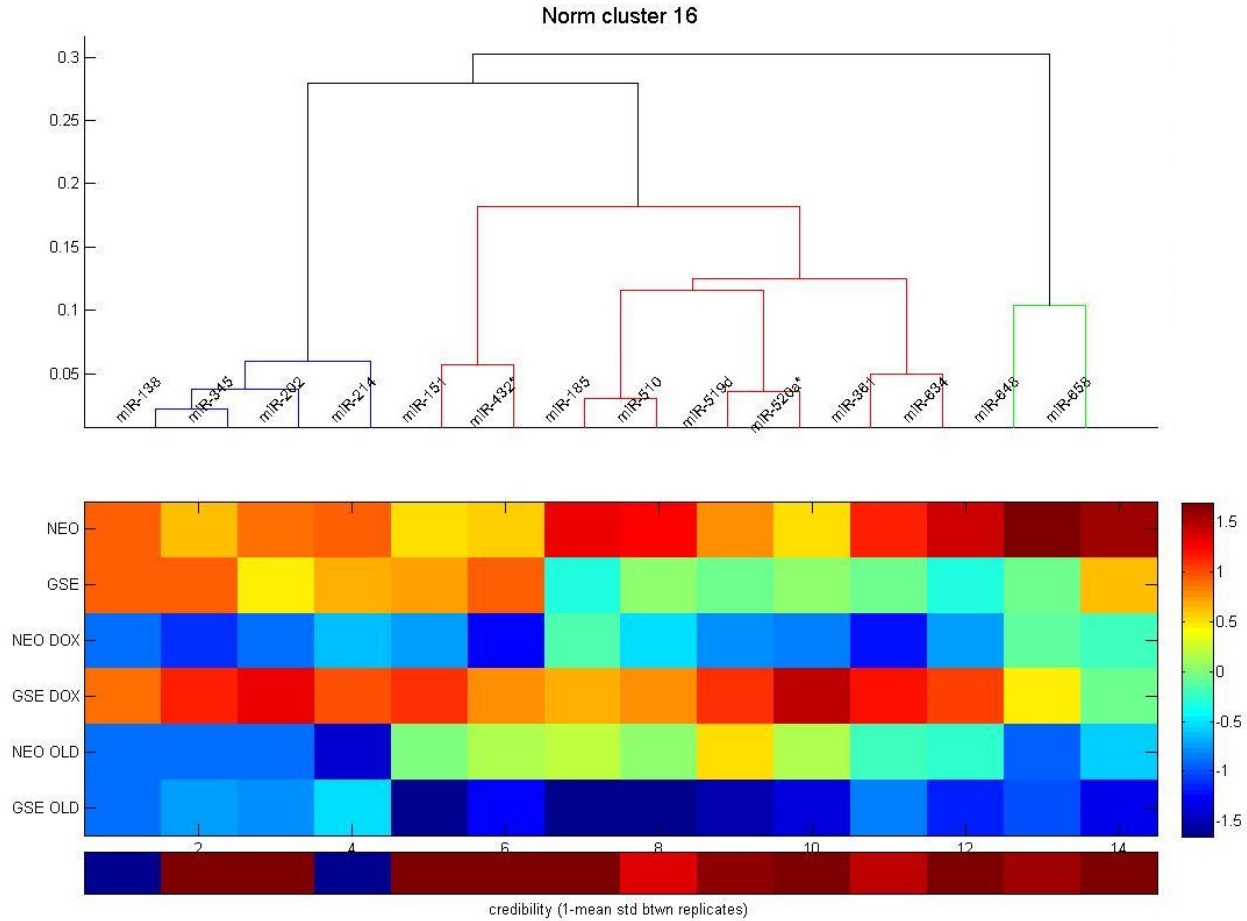


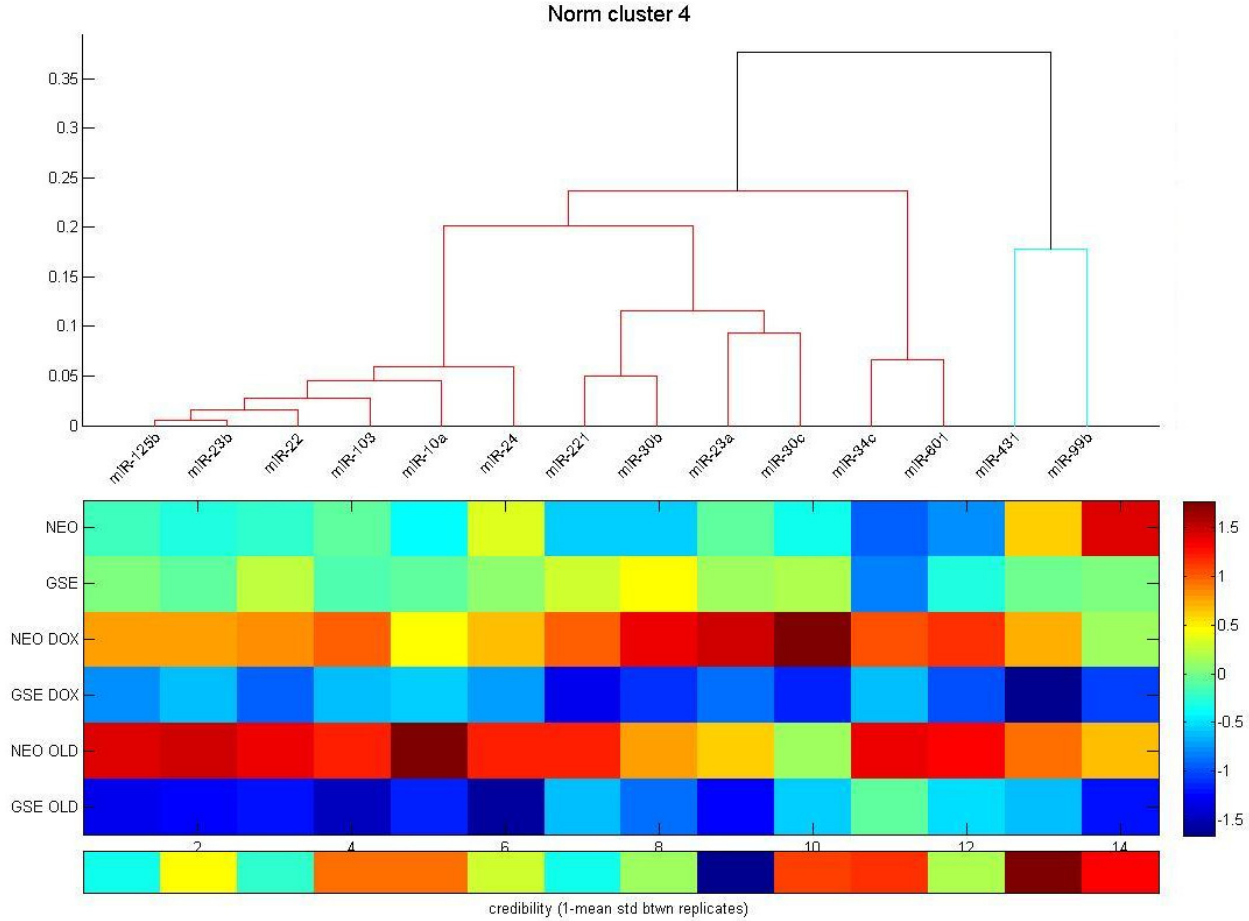

Norm cluster 13
0.3
0.25
0.2
0.15
0.1
0.05
0
mIR-376a*
mIR-552
mIR-127
mIR-493-3p
mIR-495
mIR-368
mIR-432
mIR-557
mIR-136
mIR-98
mIR-148b
mIR-27b
mIR-27a
mIR-29c
mIR-134
mIR-385
mIR-299-5p
mIR-30d
mIR-210
mIR-516-5p
mIR-302c*
mIR-410
mIR-331
mIR-329
mIR-382
mIR-802
mIR-491
R-133a-133b
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
5
10
15
20
25
2
1.5
1
0.5
0
-0.5
-1
credibility (1-mean std btwn replicates)

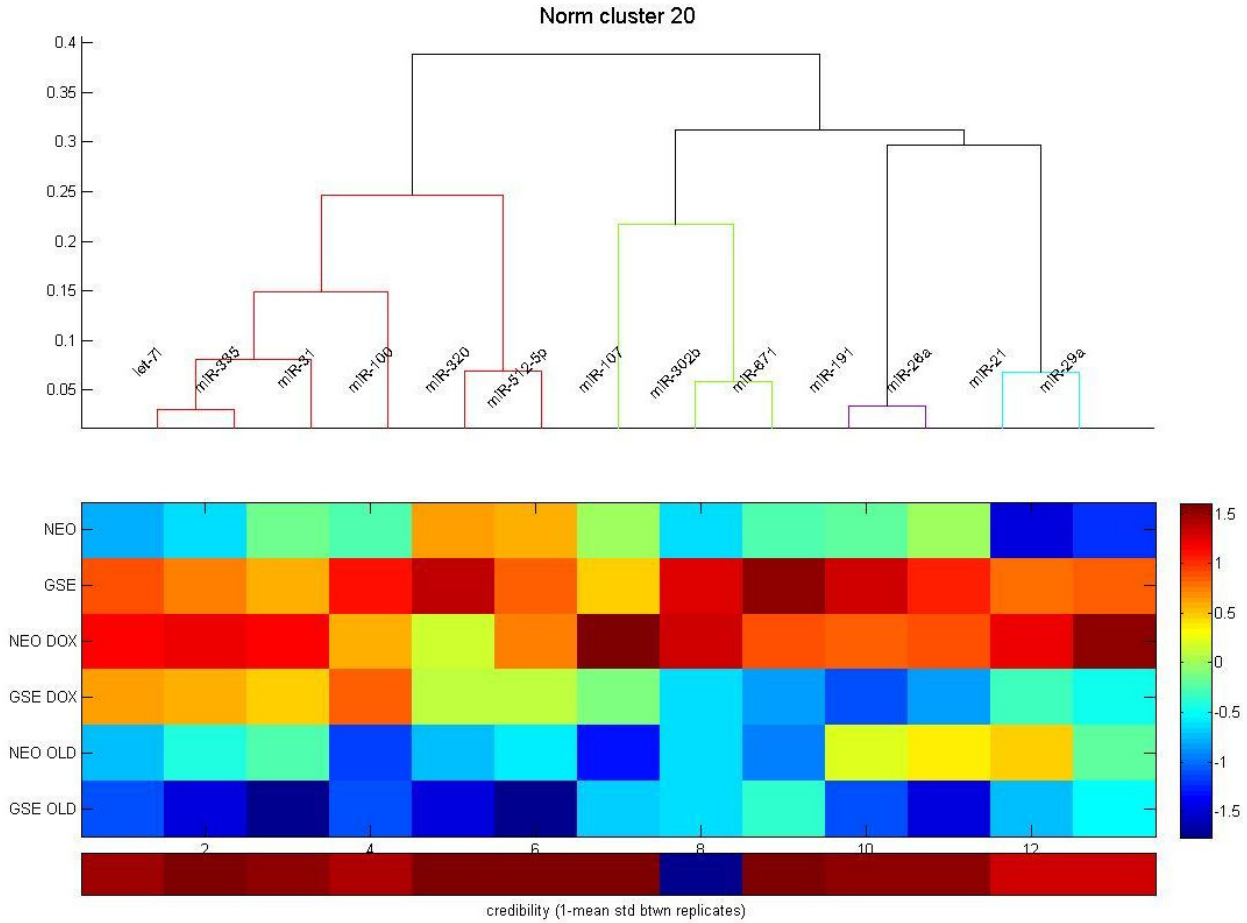

Norm cluster 15
0.3
0.25
0.2
0.15
0.1
0.05
0
mIR-211
mIR-653
mIR-25
mIR-627
mIR-199b
mIR-148a
mIR-181b
mIR-487a
mIR-574
mIR-299-3p
mIR-28
mIR-485-3p
mIR-324-5p
mIR-381
mIR-208
mIR-638
let-7e
mIR-616
mIR-645
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
14
16
18
1.5
1
0.5
0
-0.5
-1
-1.5
-2
credibility (1-mean std btwn replicates)

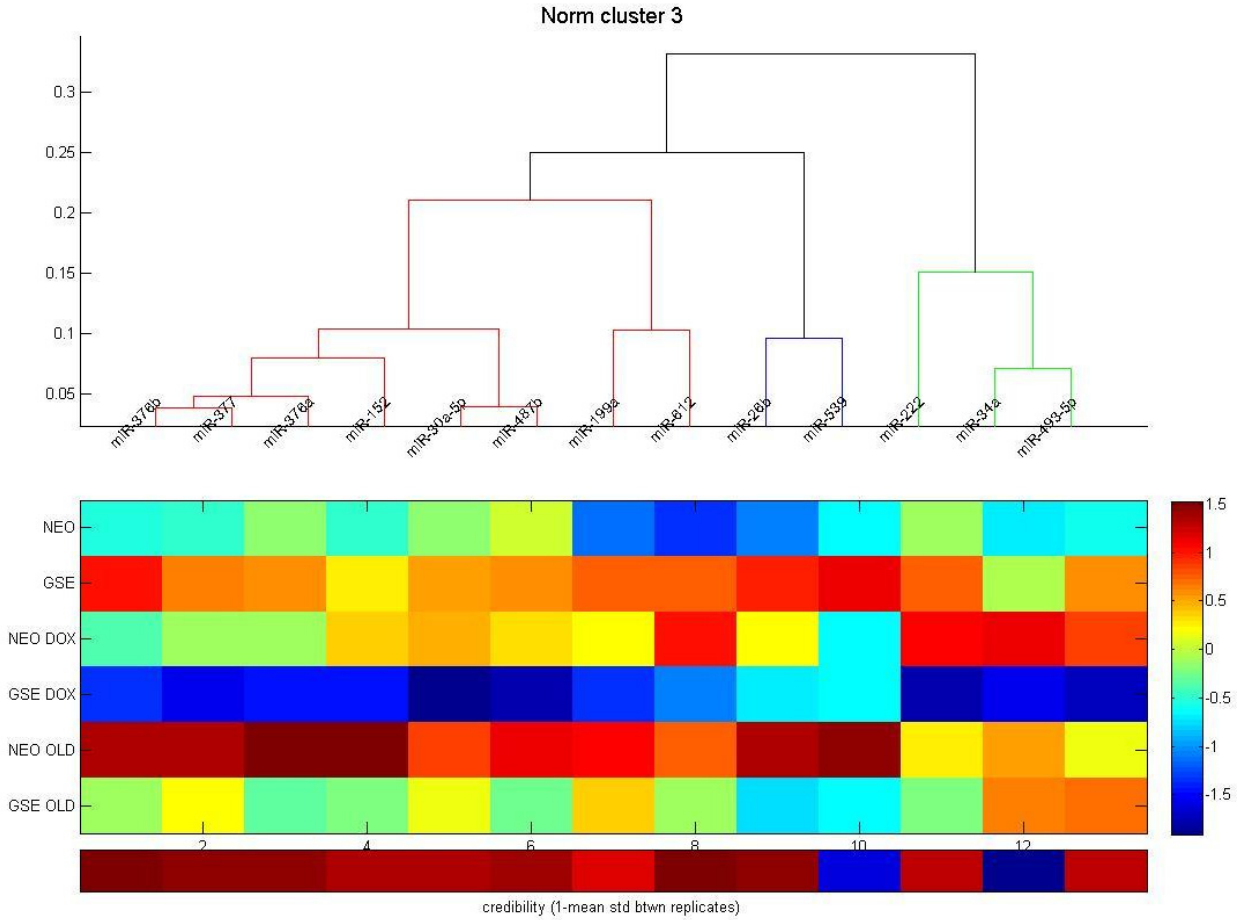

Norm cluster 12
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
miR-195
miR-424
miR-125a
miR-15a
let-7b
let-7c
miR-542-5p
let-7g
miR-494
miR-513
miR-92b
miR-551a
miR-769-3p
miR-503
miR-584
miR-652
miR-654
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
14
16
1.5
1
0.5
0
-0.5
-1
-1.5
credibility (1-mean std btwn replicates)


Norm cluster 19
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
miR-489
miR-595
miR-193a
miR-193b
miR-498
miR-199a*
miR-572
miR-663
miR-298
miR-518b
miR-29b
miR-492
miR-30e-5p
miR-373*
miR-384
miR-497
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
14
16
1.5
1
0.5
0
-0.5
-1
-1.5
credibility (1-mean std btwn replicates)


Norm cluster 16
0.3
0.25
0.2
0.15
0.1
0.05
miR-138
miR-345
miR-292
miR-214
miR-151
miR-432*
miR-185
miR-510
miR-519d
miR-520a*
miR-381
miR-634
miR-648
miR-658
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
14
1.5
1
0.5
0
-0.5
-1
-1.5
credibility (1-mean std btwn replicates)


Norm cluster 4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
miR-125b
miR-23b
miR-22
miR-103
miR-10a
miR-24
miR-221
miR-30b
miR-23a
miR-30c
miR-34c
miR-801
miR-431
miR-99b
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
14
1.5
1
0.5
0
-0.5
-1
-1.5
credibility (1-mean std btwn replicates)


Norm cluster 20
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
let-7i
mIR-395
mIR-31
mIR-100
mIR-320
mIR-512-5p
mIR-107
mIR-302b
mIR-871
mIR-191
mIR-26a
mIR-21
mIR-29a
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
1.5
1
0.5
0
-0.5
-1
-1.5
credibility (1-mean std btwn replicates)


Norm cluster 3
0.3
0.25
0.2
0.15
0.1
0.05
miR-376b
miR-377
miR-376a
miR-152
miR-30a-5p
miR-487b
miR-199a
miR-612
miR-26b
miR-539
miR-222
miR-34a
miR-493-5p
NEO
GSE
NEO DOX
GSE DOX
NEO OLD
GSE OLD
2
4
6
8
10
12
1.5
1
0.5
0
-0.5
-1
-1.5
credibility (1-mean std btwn replicates)

Cell Death and Differentiation (2009), 1–10

www.nature.com/cdd

p53-independent upregulation of miR-34a during oncogene-induced senescence represses MYC

NR Christoffersen[1], R Shalgi[2,3,6], LB Frankel[1,6], E Leucci[1], M Lees[1], M Klausen[4,7], Y Pilpel[3], FC Nielsen[4], M Oren[2] and AH Lund*[,1,5]

Aberrant oncogene activation induces cellular senescence, an irreversible growth arrest that acts as a barrier against tumorigenesis. To identify microRNAs (miRNAs) involved in oncogene-induced senescence, we examined the expression of miRNAs in primary human TIG3 fibroblasts after constitutive activation of B-RAF. Among the regulated miRNAs, both miR-34a and miR-146a were strongly induced during senescence. Although members of the miR-34 family are known to be transcriptionally regulated by p53, we find that miR-34a is regulated independently of p53 during oncogene-induced senescence. Instead, upregulation of miR-34a is mediated by the ETS family transcription factor, ELK1. During senescence, miR-34a targets the important proto-oncogene MYC and our data suggest that miR-34a thereby coordinately controls a set of cell cycle regulators. Hence, in addition to its integration in the p53 pathway, we show that alternative cancer-related pathways regulate miR-34a, emphasising its significance as a tumour suppressor.
Cell Death and Differentiation advance online publication, 21 August 2009; doi:10.1038/cdd.2009.109

Senescence is a cellular stress response that results in a permanent proliferative arrest of the cell. The term cellular senescence was originally used to describe the limited proliferative potential of cultured cells, caused by the gradual shortening of telomeres with each round of replication, leading to the activation of the DNA double-strand break checkpoint.[1–3] During the last decade it has become evident that cellular senescence can occur in a number of situations that do not involve telomere dysfunction, including DNA-replicatory stress, oncogene activation and oxidative stress.

Oncogene-induced senescence was first reported as a result of constitutive activation of RAS in primary cells.[4] After an initial proliferative burst, cells become growth arrested and display morphological changes associated with cellular senescence, accompanied by the upregulation of $p14^{ARF}$/$p19^{ARF}$ and $p16^{INK4a}$.[5] The physiological relevance of this phenomenon has been questioned, but recently, *in vivo* induction of senescence has been shown for a number of oncogenes,[6,7] including B-RAF in human benign tumours and N-RAS, K-RAS, H-RAS, B-RAF and E2F3 in mouse models. *In vivo* senescence has also been reported as a consequence of tumour suppressor inactivation for PTEN and NF1 in human benign tumours, and it has been shown that restoration of functional p53 leads to senescence, which prevents tumour progression *in vivo*.[6,7] From these reports, it is evident that oncogene-induced senescence constitutes a barrier against tumorigenesis. Oncogenes that elicit a senescence response often converge on the activation of p53 and/or RB,[7–9] although RAF-induced senescence independent of both p53 and RB has been reported in human cells.[10] The capacity of different oncogenes to trigger senescence seems to depend on cell and tissue type, perhaps reflecting the integrity of tumour suppressor networks. In human fibroblasts, inactivation of neither p53 nor RB alone overcomes RAS-induced senescence, whereas simultaneous inactivation of both pathways does.[4] It is therefore important to identify additional molecular mechanisms involved in senescence.

MicroRNAs (miRNAs) are important regulators of gene expression and they are involved in virtually all cellular processes, including proliferation, differentiation, metastasis and apoptosis.[11] Accordingly, it has been shown that some miRNAs may be categorised as *bona fide* tumour suppressors or proto-oncogenes.[12] In this study, we investigate the involvement of miRNAs in oncogene-induced senescence. To address this issue, we identified differentially regulated miRNAs in human diploid fibroblasts undergoing oncogene-induced senescence. We report a strong upregulation of miR-34a during B-RAF-induced senescence. The miR-34 family members (a, b and c) have attracted much attention because of their identification as p53 target genes and their reported involvement in p53-mediated processes, such as cell cycle arrest and apoptosis.[13–18] We show that miR-34a upregulation during B-RAF-induced senescence is independent of p53. Rather, the regulation is mediated by ELK1, a previously unreported regulator of miR-34a transcription belonging to the ETS family of transcription factors. Furthermore, we find that

[1]Biotech Research & Innovation Centre, University of Copenhagen, Copenhagen, Denmark; [2]Department of Molecular Cell Biology, the Weizmann Institute, Rehovot, Israel; [3]Department of Molecular Genetics, the Weizmann Institute, Rehovot, Israel; [4]Department of Clinical Biochemistry, Copenhagen University Hospital, Copenhagen, Denmark and [5]Center for Epigenetics, University of Copenhagen, Copenhagen, Denmark
*Corresponding author: AH Lund, Biotech Research & Innovation Centre, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark.
Tel: +45 353 256 57; Fax: +45 353 256 69; E-mail: anders.lund@bric.dk
[6]These authors contributed equally to this work.
[7]Current address: Novozymes A/S, Krogshoejvej 36, DK-2880 Bagsvaerd, Denmark
Keywords: miRNA; miR-34a; B-RAF; senescence; MYC
Abbreviations: 4-OHT, 4-hydroxytamoxifen; 4-tU, 4-thiouridine; 5-FU, 5-fluorouracil; ChIP, chromatin immunoprecipitation; EtOH, ethanol; miRNA, microRNA; PI, propidium iodide; SA-β-gal, senescence-associated beta-galactosidase
Received 17.2.09; revised 07.7.09; accepted 09.7.09; Edited by G Melino

miR-34a targets the important proto-oncogene MYC during B-RAF-induced senescence, suggesting that miR-34a through MYC repression mediates indirect downregulation of an entire set of mitotic genes during B-RAF-induced senescence.

Results

B-RAF oncogene activation regulates miRNAs. To identify miRNAs involved in oncogene-induced senescence, we examined the expression of miRNAs in hTERT-immortalised TIG3 TERT/ΔB-RAF:ER cells stably expressing a conditional B-RAF construct, ΔB-RAF:ER.[19] In human fibroblasts, ΔRAF-1:ER induces irreversible cellular senescence through activation of the MAP-kinase pathway, which is accompanied by increased levels of $p16^{INK4a}$ but does not depend on p53 and p21.[20] In accordance with a senescent phenotype, TIG3 TERT/ΔB-RAF:ER cells become senescent within 3–5 days of B-RAF activation (through treatment with 500 nM 4-hydroxytamoxifen, 4-OHT), as assessed by morphological changes, stalled growth, increased senescence-associated beta-galactosidase (SA-β-gal) activity, formation of senescence-associated heterochromatic foci (data not shown) and gradually increased expression of $p16^{INK4a}$ (Supplementary Figure S1).

We arrayed miRNA expression in normal *versus* senescent TIG3 TERT/ΔB-RAF:ER cells after 3 days of B-RAF activation. Eighteen miRNAs were significantly regulated across four biological replicates ($P<0.001$, Table 1). The results were validated for miR-146a and miR-34a that were most prominently regulated (Figure 1). The absolute levels of miR-146a were low (Supplementary Figure S2), which prompted us to focus on miR-34a, which was approximately eightfold upregulated after 3 days of B-RAF activation (Table 1). The family members miR-34b and -34c were not detected on the arrays and their lack of expression was validated by quantitative PCR (qPCR) (data not shown).

Table 1 Differentially expressed miRNAs in TIG3 TERT/ΔB-RAF:ER cells undergoing senescence (OIS) *versus* normal cells

	Unique ID	OIS/normal	*P*-value
1	hsa-miR-146a	11.27	<0.0001
2	hsa-miR-34a	8.85	<0.0001
3	hsa-miR-29b	3.23	<0.0001
4	hsa-miR-31	3.09	<0.0001
5	hsa-miR-154*	3.03	<0.0001
6	hsa-miR-532	2.93	<0.0001
7	hsa-miR-376b	2.91	0.0001
8	hsa-miR-132	2.76	0.0001
9	hsa-miR-376a	2.69	<0.0001
10	hsa-miR-425-5p	2.22	0.0005
11	hsa-miR-495	2.07	0.0008
12	hsa-miR-660	2.05	0.0009
13	hsa-miR-27b	0.33	<0.0001
14	hsa-miR-193b	0.34	0.0001
15	hsa-miR-335	0.45	<0.0001
16	hsa-miR-765	0.56	0.0005
17	hsa-miR-30a-3p	0.59	0.0001
18	hsa-miR-421	0.71	0.0006

*Shown is the mean fold change of four biological replicates and the corresponding *P*-value

Cellular effects of miR-34a. To characterise the cellular effects of miR-34a, we transfected TIG3 TERT/ΔB-RAF:ER cells with a miR-34a precursor and analysed the effect on cellular morphology, growth and cell cycle progression. Cells overexpressing miR-34a exhibited a senescence-like morphology (Supplementary Figure S3). In accordance with previous studies in primary cell cultures,[17] overexpression of miR-34a reduced cellular proliferation (Figure 2a), resulting from an accumulation of cells in the G1 phase of the cell cycle and concomitant reductions of the cell populations in S and G2/M phases (Figure 2b).

p53-independent regulation of miR-34a. Members of the miR-34 family are direct transcriptional targets of p53, and the miR-34 gene promoters contain p53-binding sites that are conserved among humans and rodents.[13,14,16–18] Several studies have shown p53-dependent upregulation of miR-34a in human and mouse cells, as well as in mouse models, as a consequence of DNA damage.[13,14,16–18,21] To assess the importance of p53 for miR-34a regulation during B-RAF-induced senescence, we depleted p53 in TIG3 TERT/ΔB-RAF:ER cells using siRNA. Surprisingly, although p53 depletion resulted in a moderate decrease in the level of miR-34a in normal cells, it had little effect on the degree of miR-34a upregulation after B-RAF activation (Figure 3a). Efficient p53 knockdown was verified by western blotting (Figure 3b). In addition, p53 knockdown prevented the induction of p53 and its target gene, p21, after treatment of the cells with the DNA damage-inducing agent 5-fluorouracil (5-FU) (Supplementary Figure S4a). To further verify the p53-independent regulation of miR-34a during B-RAF-induced senescence, we produced TIG3 TERT/ΔB-RAF:ER cells with stable expression of p53DD, a dominant-negative

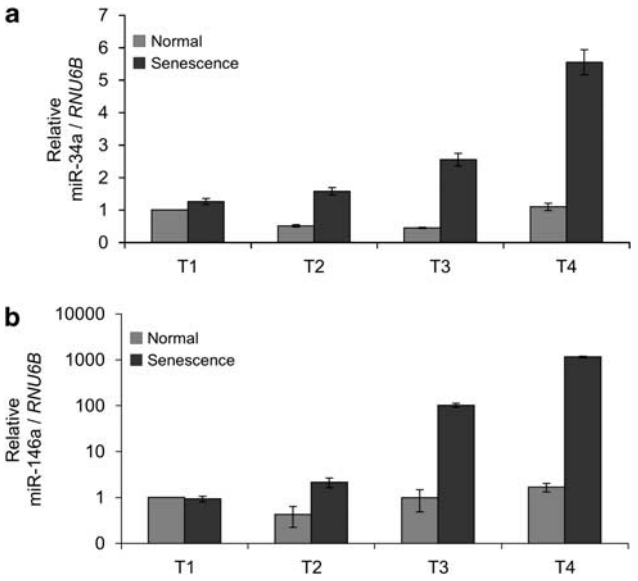


Figure 1 Validation of miRNA microarray data by qPCR. (**a** and **b**) Regulation of miR-34a and miR-146a in TIG3 TERT/B-RAF:ER cells on days 1–4 (T1–T4) after inducing senescence by B-RAF activation. Values are normalised to *RNU6B* levels and shown relative to T1 in normal cells. Data are shown as the mean ± S.D. of three replicates

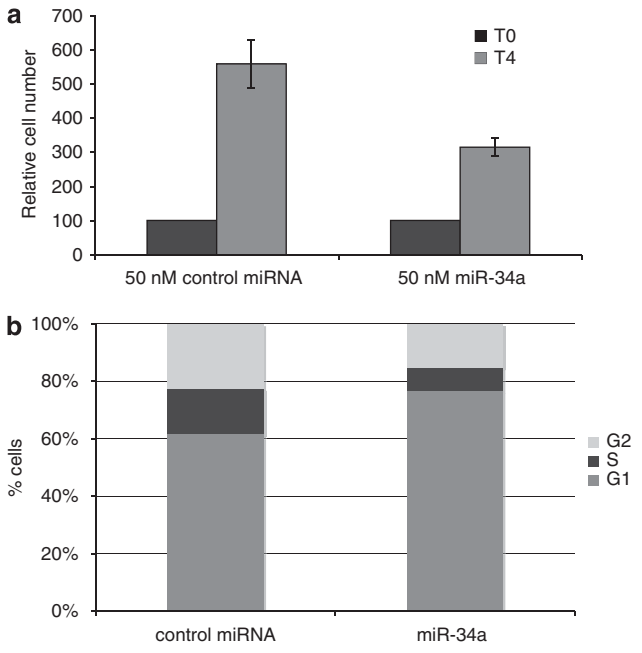


Figure 2 miR-34a inhibits growth of TIG3 TERT/B-RAF:ER cells. (**a**) Four-day growth assay of cells transfected with 50 nM miR-34a or a control miRNA. Cell number is shown relative to T0. Data are shown as the mean ± S.D. of three replicates and is representative of three independent experiments. (**b**) miR-34a affects cell cycle in TIG3 TERT/B-RAF:ER cells. Cell cycle analysis of cells 3 days after transfection with 50 nM miR-34a or a control miRNA. A representative experiment is shown

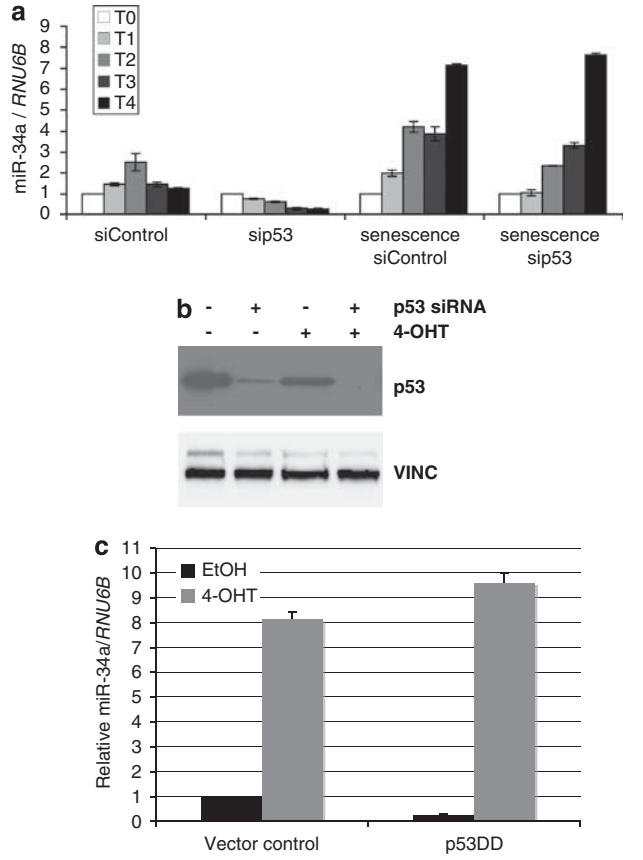


Figure 3 p53-independent increase of miR-34a expression during B-RAF-induced senescence. (**a**) miR-34a induction by B-RAF persists despite p53 knockdown. Transfection of TIG3 TERT/B-RAF:ER cells with 50 nM siRNA followed by B-RAF induction by 4-OHT treatment. Values are normalised to *RNU6B*, related to T0 and shown as the mean ± S.D. of three replicates. (**b**) Western blotting for p53 in TIG3 TERT/B-RAF:ER cells transfected with 50 nM siRNA against p53 or a control siRNA, followed by 3 days of treatment with 4-OHT. (**c**) miR-34a induction by B-RAF is intact in TIG3 TERT/B-RAF:ER cells expressing a dominant-negative variant of p53 (p53DD) on 3 days of treatment with 4-OHT

variant of the p53 gene.[22] Though the basal level of miR-34a was reduced in p53DD-expressing cells relative to the control cells expressing an empty vector, B-RAF induction in p53DD cells resulted in approximately ninefold upregulation of miR-34a relative to the level in normal p53DD cells (Figure 3c). The p53DD cells did not possess functional wild-type p53 activity, as p53 failed to induce p21 after treatment with 5-FU (Supplementary Figure S4b). Thus, we infer that miR-34a induction in this model system is dependent on mechanisms other than p53.

miR-34a expression is regulated by ELK1. The observation that regulation of miR-34a does not depend on p53 prompted us to look for alternative regulators of miR-34a expression during B-RAF-induced senescence. Phosphorylated ERK can activate a number of transcription factors, including ETS, AP-1, MYC and CREB.[23] The primary miR-34a transcript is produced by the splicing of two exons located 30 kb apart, and the area downstream of the transcription start site is highly conserved between human, mouse and rat.[16,17] Regulatory Vista (rVista)[24] analysis of this region revealed the presence of multiple conserved transcription factor-binding sites, including a number of putative binding sites for members of the ETS family (Figure 4a). To test their functional significance, we depleted individual ETS family members from TIG3 TERT/ΔB-RAF:ER cells using siRNA and measured miR-34a regulation after B-RAF activation. Although knockdown of ETS1, ETS2 and ELF1 had little effect (Supplementary Figure S5), depletion of ELK1 significantly impaired B-RAF-mediated induction of miR-34a (Figure 4b) ($P<0.04$, Student's *t*-test). Efficient knockdown of ELK1 and ELF1 was confirmed at the mRNA and protein level (Supplementary Figure S6).

To verify the ability of ELK1 to regulate the human miR-34a promoter, we cloned a 760-bp fragment of the proximal promoter containing several putative ELK1-binding sites into a luciferase reporter vector (pProm34a). Co-transfections of pProm34a with an ELK1 expression construct into HEK293 cells resulted in a significant upregulation of luciferase activity ($P<0.002$, Student's *t*-test), showing the capacity of ELK1 to regulate the miR-34a promoter (Figure 4c). To confirm the binding of ELK1 to the miR-34a promoter at the endogenous level, we carried out chromatin immunoprecipitation (ChIP) in TIG3 TERT/ΔB-RAF:ER cells. We observed a strong enrichment of ELK1 on the miR-34a promoter after B-RAF activation (Figure 4d), consistent with the notion that ELK1 is an important regulator of miR-34a during oncogene-induced senescence in primary human fibroblasts.

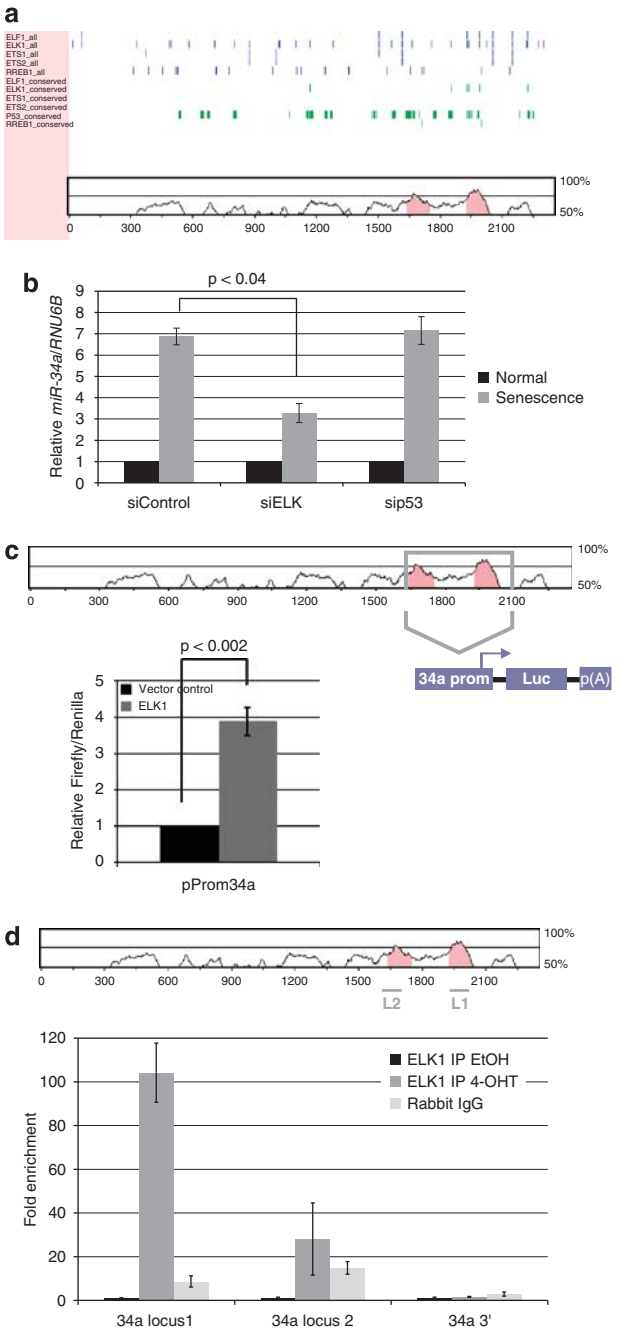


Figure 4 miR-34a expression is regulated by ELK1 in TIG3 TERT/B-RAF:ER cells undergoing senescence. (**a**) Vista sequence alignment of human and mouse miR-34a promoters with regulatory Vista analysis of transcription factor-binding motifs (Ensembl positions: hsa chr. 1:9.164.107–9.166.456 and mmu chr. 4:148.890.943–148.894.943). Blue bars: motifs present in the human promoter; green bars: conserved motifs. (**b**) B-RAF-induced upregulation of miR-34a is attenuated by knockdown of ELK1. Values are normalised to *RNU6B* and shown relative to normal cells. Data are shown as the mean ± S.E.M. of three independent experiments. (**c**) Exogenous ELK1 expression upregulates a luciferase reporter construct containing part of the human miR-34a promoter (pProm34a). Luciferase values (pProm34a) normalised to Renilla luciferase are shown relative to the empty vector control. (**d**) ELK1 binding to the miR-34a promoter increases on B-RAF activation. Chromatin immunoprecipitation of ELK1 with qPCR detection of two different loci (L1 and L2) within the miR-34a promoter as indicated in the upper panel. A downstream miR-34a locus is shown as negative control. Values are shown as the mean ± S.D. of three replicates and are representative of three independent experiments

Identification of miR-34a targets in oncogene-induced senescence. To understand the role of miR-34a in oncogene-induced senescence, we tested global mRNA expression using microarrays of TIG3 TERT/ΔB-RAF:ER cells transfected with a miR-34a LNA inhibitor or a scrambled control LNA in the presence or absence of B-RAF activation. Affymetrix microarrays (Affymetrix, Santa Clara, CA, USA) were performed with total RNA from three biological replicates for each treatment. miR-34a and scrambled inhibitors were transfected into the cells at day 2 of 4-OHT treatment (or EtOH as vector control) and the samples were harvested 24 h after transfection.

As expected, B-RAF activation resulted in major changes in gene expression, as evident from a hierarchical cluster analysis (Supplementary Figure S7 and Supplementary Table S3). Importantly, miR-34a inhibition resulted in de-repression of several transcripts that have previously been reported as miR-34a targets, including *BCL2* and *CDK6*, and, to a lesser extent, *MET* and *CCND1* (Supplementary Table S1).

To highlight changes related to miR-34a, we carried out hierarchical clustering analysis of the top 20% of the genes that were most influenced by miR-34a inhibition (Figure 5a). Although we did not detect an enrichment of motifs matching the miR-34a seed sequence within the 3′UTRs of genes in clusters resulting from this data set (~1800 transcripts), several clusters displayed an interesting miR-34a dependency. We focused on a cluster of ~350 transcripts, which were repressed upon B-RAF activation. Interestingly, the inhibition of miR-34a alleviated B-RAF-mediated repression of this cluster, which we have termed the 'B-RAF-repressed, miR-34a-influenced cluster' (Figure 5a). Functional annotation analysis[25] of the cluster revealed a significant enrichment for genes related to the M phase of the cell cycle ($P<5.7\times10^{-15}$), the cell cycle ($P<1.9\times10^{-11}$) and other related functions (Supplementary Table S2). Furthermore, *CDK6* and *BCL2*, two well-known targets of miR-34a,[13,26] were included in this cluster. As most of the transcripts in this cluster did not contain binding sites for miR-34a in their 3′UTRs, we speculated that a common transcriptional regulator, which is a miR-34a target, could be responsible for the observed expression pattern and mediate a global miR-34a effect. Interestingly, when subjecting the promoters of the genes in the 'B-RAF-repressed, miR-34a-influenced cluster' to motif finding (using AMADEUS[27]), a marked enrichment for a motif resembling a MYC-binding site was found ($P<3.3\times10^{-12}$), suggesting that miR-34a could influence gene expression through targeting of MYC.

miR-34a targets MYC during oncogene-induced senescence. Inspection of the 3′UTR sequence of *MYC* revealed the presence of a perfectly complementary and evolutionarily conserved 7-nucleotide match to the seed region of miR-34b and miR-34c, and a 6-nucleotide seed match to miR-34a (Supplementary Figure S8). miR-34b and miR-34c were recently reported to regulate MYC,[26,27] but neither miR-34b nor miR-34c was detected above background in our miRNA microarray experiments on TIG3 TERT/ΔB-RAF:ER cells.

In the Affymetrix array data, we did not detect changes in the *MYC* mRNA level after miR-34a inhibition (Supplementary

Figure S9a). Furthermore, overexpression of a miR-34a precursor in TIG3 TERT/ΔB-RAF:ER cells had little effect on *MYC* at the mRNA level (Supplementary Figure S9b), suggesting post-transcriptional regulation at the level of translation. To investigate whether miR-34a can translationally regulate MYC, we transfected TIG3 TERT/ΔB-RAF:ER cells with a miR-34a precursor and controls and analysed the effect on MYC protein using western blotting (Figure 5b). Strikingly, miR-34a overexpression resulted in marked downregulation of endogenous MYC protein, indicating that miR-34a can affect cell proliferation through repression of MYC. Similar results were found in H1299 cells (Supplementary Figure S10). To test whether miR-34a represses MYC downstream of B-RAF activation, we induced B-RAF in TIG3 TERT/ΔB-RAF:ER cells and transfected the cells 2 days later with a miR-34a LNA inhibitor or scrambled LNA. As evident from Figure 5c, inhibition of miR-34a during senescence resulted in de-repression of MYC. This strongly suggests that miR-34a can affect a large cohort of cell cycle regulators through translational repression of MYC. As we have shown that miR-34a is regulated by ELK1 downstream of B-RAF induction, we tested the ability of an siRNA against *ELK1* to phenocopy the effect of miR-34a inhibition on MYC protein. As evident from Figure 5d, depletion of ELK1 did indeed cause de-repression of MYC protein during B-RAF-induced senescence, underlining the importance of ELK1 in miR-34a regulation during oncogene-induced senescence.

To test whether miR-34a regulates MYC in a direct manner, we measured the effect of the miR-34a precursor on a luciferase reporter construct containing the *MYC* 3′UTR (pLSV-M3′). Although miR-34a significantly repressed luciferase activity of pLSV-M3′ ($P<0{,}03$, Student's *t*-test), it did not affect a mutated version (pLSV-M3′MUT) in which the miR-34a seed-binding site had been altered (Figure 5e). To further show the direct interaction between miR-34a and *MYC* mRNA, we carried out miRNA pull-out assays in which biotinylated miRNA mimics are transfected into cells, allowing for subsequent streptavidin-based purification of the mature miR-34a along with associated RNA species.[28] The mature strand furthermore contained photosensitive 4-thiouridine (4-tU) modified nucleotides, which form RNA–RNA cross-links upon long-wave UV-irradiation. miR-34a hairpins were transfected into TIG3 TERT/ΔB-RAF:ER cells in the presence and absence of B-RAF activation. The cells were UV treated to induce RNA cross-linking and after streptavidin pullout of biotinylated miR-34a, *MYC* mRNA was highly enriched in the RNA pool associated with miR-34a compared with a control mRNA, *HPRT* (Figure 5f). Thus, our data show that miR-34a regulates *MYC* through direct binding to its 3′UTR.

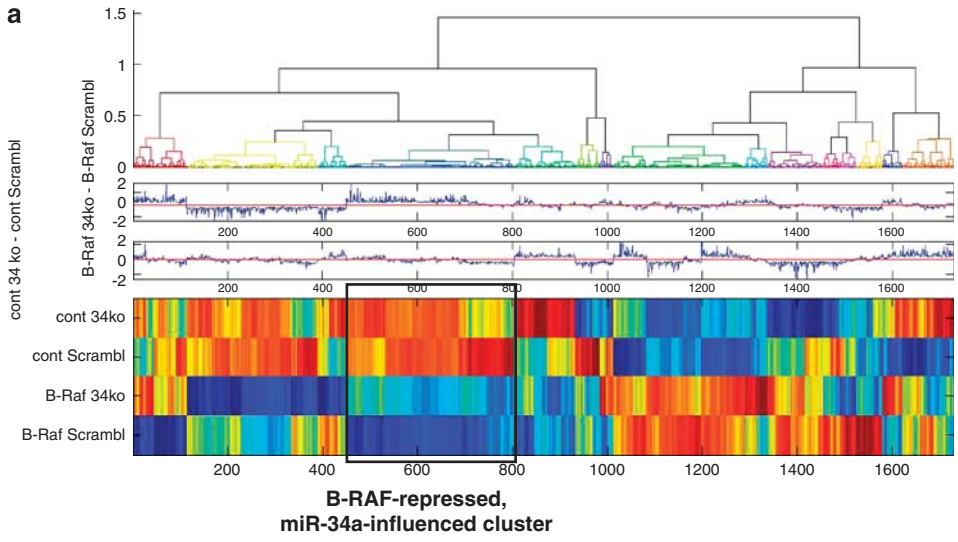
Discussion

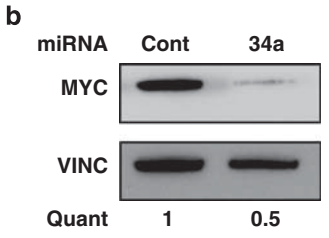
Oncogene-induced senescence is an important barrier towards cancer *in vivo*, but the underlying mechanisms are still not clear and likely differ between cell types. miRNAs regulate numerous cellular processes and a current challenge is to understand their functions and incorporate individual miRNAs into cellular pathways. Here, we provide evidence that miR-34a is transcriptionally upregulated by ELK1 downstream of B-RAF oncogene activation, which leads to senescence in human fibroblasts. This is a novel pathway that is independent of p53, which was previously reported to transcriptionally activate miR-34a. In addition, we identify MYC as a miR-34a target. In agreement with the proliferative role of MYC, we show that MYC is repressed during oncogene-induced senescence and our data suggest that this repression is mediated, at least in part, through miR-34a. The effect of miR-34a extends further to repress a set of mitotic genes that are transcriptional targets of MYC, providing a novel link between B-RAF oncogene activation and the mechanism of senescence.

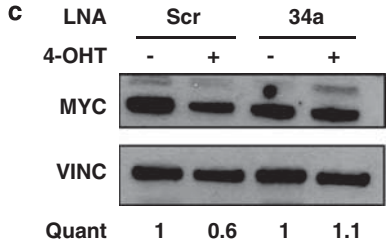
ELK1 mediates upregulation of miR-34a during B-RAF-induced senescence independently of p53. It was previously reported that exogenous expression of miR-34a induces senescence-like changes in primary cells and cancer cell lines.[17,21] Accordingly, we find that exogenous expression of a miR-34a precursor induces a partial G1 cell cycle arrest and a senescence-like cell morphology in TIG3 TERT cells. Thus, our finding that B-RAF activation induces an upregulation of miR-34a in TIG3 cells suggests that miR-34a is a mediator of B-RAF-induced senescence. Dysregulation of miR-34a has been reported for several types of cancer, suggesting its importance as a tumour suppressor. miR-34a expression is downregulated in neuroblastomas[29] and frequently reduced in pancreatic cancers cell lines.[14] Furthermore, miR-34a resides in a locus (1p36), which is frequently lost in cancer[30] and is subject to silencing because of aberrant CpG methylation of the promoter in prostate cancer and melanoma, as well as a number of cancer cell lines.[31]

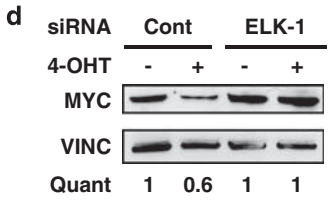
Transcriptional regulation of miR-34a has so far been ascribed to p53 and there is substantial evidence that p53 transcriptionally activates miR-34a after DNA damage, whereas the absence of p53 activity abrogates miR-34a regulation.[16,17] In contrast, we find that depletion of functional p53 by siRNA or by overexpression of dominant-negative p53DD does not markedly affect the induction of miR-34a in normal *versus* oncogene-induced senescent cells, and thus we speculate that B-RAF oncogenic stress induces miR-34a expression through different pathways than those induced by DNA damage. In support of p53-independent regulation of miR-34a downstream of B-RAF activation, it has previously been shown in human IMR-90 fibroblasts that ΔRAF-1:ER induces irreversible cellular senescence through activation of the MAP-kinase pathway independently of p53 and p21.[20] Although p53 is not necessary for miR-34a regulation in senescent TIG3 cells, p53 depletion does decrease the basal level of miR-34a in normal TIG3 cells, indicating that p53 may regulate miR-34a in this cell type under different conditions.

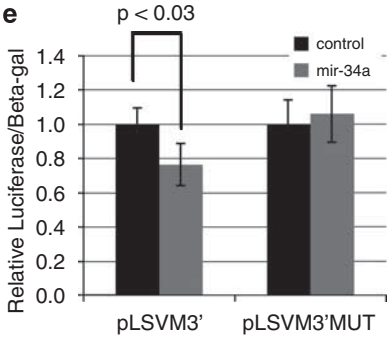
To identify novel regulators of miR-34a downstream of B-RAF, we analysed the promoter region of the human miR-34a gene for relevant transcription factor-binding sites. Among other transcription factors, B-RAF activates members of the ETS family, which has previously been implicated in cellular senescence because of the ability of ETS1 and ETS2 to activate the p16INK4a promoter.[32] The human miR-34a promoter region contains several ETS-binding motifs, including that of ETS1, ETS2, ELF1 and ELK1. It can be noted that several conserved ELK1 motifs occur in a part of the miR-34a

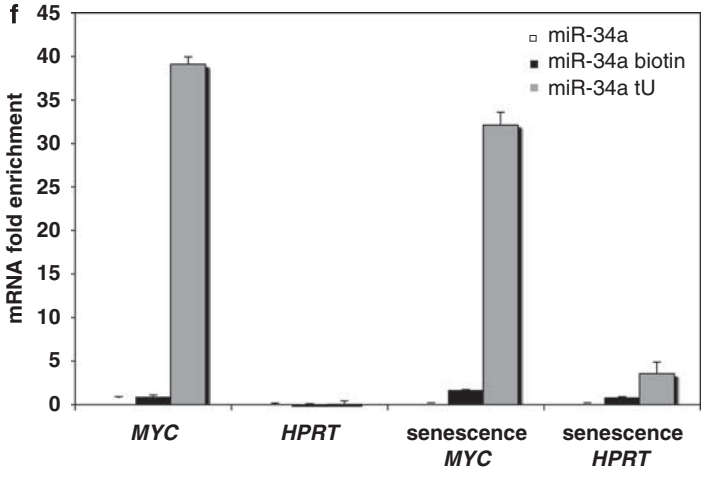

a
cont 34 ko - cont Scrambl
B-Raf 34ko - B-Raf Scrambl
cont 34ko
cont Scrambl
B-Raf 34ko
B-Raf Scrambl
B-RAF-repressed,
miR-34a-influenced cluster


b
miRNA
Cont
34a
MYC
VINC
Quant
1
0.5


c
LNA
Scr
34a
4-OHT
MYC
VINC
Quant
1
0.6
1
1.1


d
siRNA
Cont
ELK-1
4-OHT
MYC
VINC
Quant
1
0.6
1
1


e
p < 0.03
control
mir-34a
Relative Luciferase/Beta-gal
pLSVM3'
pLSVM3'MUT


f
miR-34a
miR-34a biotin
miR-34a tU
mRNA fold enrichment
MYC
HPRT
senescence
MYC
senescence
HPRT

promoter region that is highly conserved between mouse and human. Knockdown of ELK1, but not ETS1, ETS2 and ELF1, reduces the induction of miR-34a in senescent cells after B-RAF activation. Furthermore, an ELK1 expression construct markedly increases activity of a reporter construct containing part of the human miR-34a promoter. Finally, we show increased recruitment of ELK1 to the miR-34a promoter in TIG3 cells 3 days after B-RAF activation relative to control TIG3 cells. Collectively, these data show that ELK1 can activate the miR-34a promoter after B-RAF activation in a p53-independent manner.

miR-34a targets the MYC proto-oncogene during B-RAF-induced senescence. The data presented by us and others suggest a functional importance of miR-34a in senescence.[17,21,33] To identify miR-34a target genes involved in senescence, we used Affymetrix microarray analysis to measure gene expression in normal and senescent TIG3 cells treated with a miR-34a-specific inhibitor. We and others have previously used this strategy to identify targets for miR-21.[34,35] Using this approach, transcripts that are de-repressed by miR-34a in the level of mRNA degradation can be identified by their increased abundance, whereas mRNAs regulated exclusively at the level of translation cannot be directly detected. Importantly, as evident from this study, detailed analysis of the promoters of cohorts of deregulated genes can lead to the identification of translationally regulated targets. Although B-RAF activation has a major impact on gene expression, we were able to detect genes, the expression of which was alleviated by miR-34a inhibition, as visualised by hierarchical clustering of the 20% genes that are most highly regulated by the miR-34a inhibitor in normal and/or senescent TIG3 cells. We do not observe an overrepresentation of the miR-34a-binding site in this list of genes. However, consistent with previous reports, we find that miR-34a inhibition de-represses four validated miR-34a targets, namely, *BCL2*, *CDK6*, *CCND1* and *MET*,[13,16,17,36] although de-repression of *CCND1* and *MET* were modest at the mRNA level and did not belong to the top 20% genes most highly regulated by miR-34a inhibition. We therefore suspected translational inhibition to be a major component of miR-34a function, which prompted us to search for secondary regulatory effects present in the data, and to identify the primary regulator upstream to these.

Cluster analysis of the top 20% of genes affected by miR-34a depletion revealed a set of genes (the 'B-RAF-repressed, miR-34a-influenced cluster'), which is functionally enriched for mitotic genes and for genes harbouring putative MYC-binding sites in their promoters. We find that a miR-34a precursor represses MYC protein in TIG3 cells and other cell types, and importantly, we show that miR-34a inhibition de-represses MYC protein levels in cells undergoing B-RAF-induced senescence. Interestingly, we find that ELK1 depletion phenocopies the effect of miR-34a inhibition on MYC, thereby supporting the notion that ELK1 upregulates miR-34a during B-RAF-induced senescence and that miR-34a in turn represses MYC. Our data furthermore show direct interaction between miR-34a and the *MYC* 3′UTR, in that miR-34a can repress the activity of a wild type but not a mutated *MYC* 3′UTR reporter and *MYC* mRNA is highly enriched in the RNA pool associated with miR-34a in a miRNA pull-out assay.

Members of the miR-34 family have previously been linked to the MYC family of proto-oncogenes. In neuroblastoma, loss of the 1p36 locus, which encodes miR-34a as well as other potential tumour suppressors, correlates with *NMYC* amplification and miR-34a targets NMYC in several human neuroblastoma cell lines.[37] In addition, miR-34b and mir-34c were recently reported to target MYC[38,39] The binding site for miR-34a in the 3′UTRs of *MYC* is furthermore conserved in the genomes of humans, mice, rats, dogs and chicken. When taken together with the published data, our data thus suggest an evolutionarily conserved regulation of several members of the MYC family of proto-oncogenes by miR-34 family members.

Several lines of evidence implicate MYC in oncogene-induced senescence, and show that its repression is essential for the senescent phenotype. A recent report showed that MYC depletion in B-RAF or N-RAS overexpressing melanoma cells results in senescence-like phenotypes and that MYC overexpression repressed B-RAF-induced senescence.[40]

Figure 5 MYC is directly targeted by miR-34a during senescence. (**a**) Hierarchical clustering of the top 20% of transcripts most highly affected by miR-34a inhibition. Top panel: Hierarchical clustering tree. Colours on the tree mark different clusters. Middle panel: Gene fold change (log2 scale) induced by a miR-34a inhibitor in normal cells (cont/34ko–cont/scrambl) and in senescent cells (B-RAF/34ko – B-RAF/scrambl). Bottom panel: Normalised expression data. For visualisation purposes, data were gene normalised by subtracting the mean for each gene value and dividing it by the standard deviation. 'Cont 34ko': Cells treated with EtOH carrier and transfected with an LNA inhibitor of miR-34a, 'Cont Scrambled': Cells treated with EtOH carrier and transfected with a scrambled LNA control. 'B-RAF 34ko': Cells treated with 4-OHT and transfected with an LNA inhibitor of miR-34a. 'B-RAF Scrambled': Cells treated with 4-OHT and transfected with a scrambled control LNA. The 'B-RAF-repressed, miR-34a-influenced cluster' is indicated. (**b**) miR-34a overexpression represses MYC at the protein level. Western blot of MYC in TIG3 TERT/B-RAF:ER cells transfected with 50 nM miR-34a or a control miRNA. (**c**) Inhibition of miR-34a prevents reduction of MYC protein during senescence. Western blot analysis of MYC in TIG3 TERT/B-RAF:ER cells 24 h after transfection with 50 nM of a miR-34a inhibitor or a scrambled control LNA, in the presence or absence of B-RAF induction (3 days). Numbers indicate quantification of the MYC band densities relative to Vinculin. (**d**) Depletion of ELK1 by siRNA prevents reduction of MYC protein during senescence. Western blot analysis of MYC in TIG3 TERT/B-RAF:ER cells 24 h after transfection with 50 nM of siRNA against *ELK1* or a control siRNA, followed by 3 days of B-RAF induction. Numbers indicate quantification of the MYC band densities relative to Vinculin. (**e**) miR-34a regulates MYC through binding to the *MYC* 3′UTR. TIG3 TERT/B-RAF:ER cells were co-transfected with miR-34a or a control miRNA and a wild type or mutated version of the *MYC* 3′UTR cloned into a luciferase vector (pLSVM3′ and pLSVM3′MUT, respectively). Luciferase values were normalised to β-galactosidase activity and are shown relative to control as the mean $+/-$S.D. of three replicates. The data are representative of three independent experiments. (**f**) miRNA pull-out assay demonstrating direct binding of miR-34a to the 3′UTR of *MYC* mRNA in normal and senescent TIG3 TERT/B-RAF:ER cells. After 2 days of treatment with 4-OHT or EtOH carrier, TIG3 TERT/B-RAF:ER cells were transfected with 30 nM of an unmodified miR-34a duplex (miR-34a), a biotinylated miR-34a duplex (miR-34a Biotin) or a mix of two biotinylated miR-34a duplexes containing UV-reactive 4-thiouridine (4-tU) nucleotides at positions 7 and 11, respectively (miR-34a tU). The transfected cells were UV irradiated to induce cross-linking of the 4-tU nucleotides to associated mRNAs. On extraction of biotinylated duplexes, the presence of *MYC* mRNA or a control (HPRT) mRNA was measured by qPCR

Likewise, inactivation of MYC in primary tumours induced by conditional MYC overexpression induces senescence and tumour regression,[41] and reduced MYC levels in normal diploid human fibroblasts increases the frequency of telomere-independent senescence in a p16-dependent manner.[42] In view of this, we propose that miR-34a functions downstream of B-RAF to downregulate MYC protein levels in TIG3 TERT/ΔB-RAF:ER cells, thereby promoting B-RAF-induced senescence. However, in this cellular system, inhibition of miR-34a alone did not prevent the appearance of phenotypic hallmarks of senescence, such as senescence-associated heterochromatic foci and increased SA-β-gal activity (data not shown). Hence, miR-34a is an important but not a sole player in B-RAF-induced senescence in these cells, and other mediators have yet to be identified.

In summary, we have shown that human primary cells undergoing B-RAF-induced senescence strongly upregulate miR-34a in a p53-independent manner. Instead, the regulation is mediated, at least partly, by the ETS family transcription factor, ELK1. In addition, during senescence miR-34a affects the expression of a cluster of mitotic genes through translational repression of MYC. This places miR-34a at a key node, responding to several independent cancer-associated pathways, and emphasises the importance of miR-34a as a tumour suppressor.

Materials and Methods

Cell culture. TIG3 TERT/B-RAF:ER cells are primary human diploid fibroblasts immortalised by hTERT. B-RAF is constitutively active because of truncation of the regulatory N-terminal domain and is fused to the hormone-binding domain of the oestrogen receptor, which was modified to respond to 4-OHT but not β-estradiol.[43] TIG3 and HEK293 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% FBS (Biochrom AG, Berlin, Germany), 100 U/ml penicillin and 100 μg/ml streptomycin (Invitrogen, Carlsbad, CA, USA) and incubated at 37°C in 5% CO_2. For conditional activation of B-RAF, TIG3 TERT/B-RAF:ER cells were treated for 3 days with 500 nM 4-OHT or equal volumes of EtOH carrier. For 5-FU treatment, cells were incubated for 16 h with 50 μg/ml 5-FU or equal volumes of DMSO carrier.

TIG3 TERT/B-RAF:ER p53DD cells were produced by infection of TIG3 TERT/B-RAF:ER with p53DD virus followed by puromycin selection.

miRNA precursors, anti-miRNA oligonucleotides and siRNA. The miRNA precursors were purchased from Ambion (Austin, TX, USA) and LNA-modified oligonucleotide miRNA inhibitors from Exiqon (Vedbaek, Denmark). SMARTpool siRNAs were purchased from Dharmacon (Lafayette, CO, USA), except the siRNA against p53, which was an annealed duplex of the following oligonucleotides purchased from Biosynthesis (Lewiseville, TX, USA):

hsa p53i S 5′-CTACATGTGTAACAGTTCCUU-3′ and
hsa p53i AS 5′-(P)-GGAACTGTTACACATGTAGUU-3′.

The AllStars-negative control siRNA (Qiagen, Valencia, CA, USA) was used as control for transfections with miRNA precursors and siRNAs.

Vector constructs. *pProm34a*: A 760-bp fragment of the human miR-34a promoter was PCR amplified from human genomic DNA and cloned into the pGL3 Basic vector (Invitrogen) using *XhoI* and *HindIII* restriction enzymes. The primer sequences were (restriction sites are underlined):

FW 5′-CTCGAGCGAGCAGGAAGGAGGACCCG-3′ and
RV 5′-AAGCCTGGGCTCCAGCCAGCAGGG-3′.

The ELK1 expression construct was kindly provided by Dr. Robert A Hipskind. The *MYC* luciferase constructs pLSV-M3′ and pLSV-M3′MUT were kindly provided by Dr. Martin Bushell.

Reporter assays. For promoter luciferase assays, HEK293 cells were seeded at 10 000 per 96 well and transfected (Lipofectamine 2000, Invitrogen) on the next day with 150 ng pProm34a, 25 ng pRL-TK, and 100 ng of an ELK1 expression vector or an empty pcDNA3.1 + vector. At 48 h after transfection, luciferase activity was measured using the Dual-Glo luciferase assay (Promega, Stockholm, Sweden).

For *MYC* 3′UTR luciferase assays, 80 000 TIG3 TERT/B-RAF:ER cells per 24 well were reverse transfected (Lipofectamine 2000) with 1.75 μg pLSV-M3′ or pLSV-M3′MUT, 50 nM miR-34a or AllStars control, and 0.25 μg lacZ expression vector (pCMV-b-gal, Clontech, Palo Alto, CA, USA). At 20 h after transfection, cells were washed once in PBS, lysed in 100 μl Passive Lysis Buffer (Promega) and incubated for 20 min at RT. In all, 10 μl of lysate was mixed with 90 μl of complete luciferase buffer (25 mM glycylglycine, 15 mM $MgSO_4$, 4 mM EGTA, 100 mM KH_2PO_4 buffer, 1 M DTT, 0.2 M ATP (Sigma) and 10 μM Luciferin (Sigma)) and Luciferase activity was measured immediately. β-galactosidase activity was measured by mixing 25 μl lysate, 175 μl Z-buffer (100 mM Na_2PO_4 pH 7, 10 mM KCl, 1 mM Mg_2SO_4 and 5 mM DTT) and 40 μl ortho-nitrophenyl-β-galactoside (ONPG) reaction mix (4 mg/ml ONPG, 100 mM Na_2PO_4 pH 7), and incubating at 37°C. Production of ortho-nitrophenol was measured at 410 nm.

qPCR analysis. For qPCR of mRNA, TIG3 TERT/B-RAF:ER cells were seeded at 250 000 per six well, transfected twice on two successive days with 50 nM siRNA using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol, and treated with 500 nM 4-OHT or equal volumes of ethanol the next day. Total RNA was prepared using TRIzol reagent (Invitrogen), treated with DNase and reverse transcribed using TaqMan Reverse Transcription kit (Applied Biosystems, CA, USA) with random hexamer primers. qPCR detection of human *p53*, *ELF1*, *ELK1*, *MYC* and *GAPDH* was performed with TaqMan gene expression assays (Applied Biosystems) according to the manufacturer's protocol. For miRNA qPCR, total RNA was prepared using TRIzol reagent. Reverse transcription and qPCR analyses were carried out with TaqMan miRNA assays (Applied Biosystems) for hsa-miR-34a, hsa-miR-146a and *RNU6B*.

Antibodies and western blot analysis. TIG3 TERT/B-RAF:ER cells were seeded at 250 000 per six well, transfected with 50 nM miRNA precursor using Lipofectamine 2000 and treated with 500 nM 4-OHT or equal volumes of ethanol the next day. Cells were harvested, washed once in PBS and lysed in RIPA buffer (150 nM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris–HCl pH 8, 2 mM EDTA) containing 1 mM DTT and 1 mM Pefabloc (Roche, Indianapolis, IN, USA). In all, 20 μg protein/lane was separated on a 4–20% NuPAGE Bis–Tris gel (Invitrogen) and transferred to a nitrocellulose membrane. The antibodies used were as follows – ELF1: sc-631, Santa Cruz Biotechnology (Santa Cruz, CA, USA); ELK1: #9182, Cell Signaling (Denvers, MA, USA); MYC: #9402, Cell Signaling; p16: monoclonal Ab DCS50; p53: sc-126, Santa Cruz Biotechnology; Vinculin: V9131, Sigma. The antibodies for ChIP were ELK1: sc-355X, Santa Cruz Biotechnology; Rabbit IgG (Sigma I 8140).

Cell cycle analysis. To analyse the effect of miR-34a overexpression on the cell cycle, TIG3 TERT/B-RAF:ER cells were transfected with 50 nM of miR-34a or AllStars control, stained for DNA content using propidium iodide and analysed on a FACSCalibur flow cytometer (Becton-Dickinson, Franklin Lakes, NJ, USA). Briefly, the cells were harvested by trypsinisation and washed once in PBS before fixing o/n in 70% EtOH. To stain the DNA, the cells were pelleted, resuspended in 100 μl EtOH and stained for 1 h with 300 μl PI staining solution (0.05 mg/ml propidium iodide, 20 μg/ml RNase A in 0.1% BSA).

ChIP. TIG3 TERT/B-RAF:ER cells were treated with 500 nM 4-OHT or equal volumes of ethanol the day after seeding at 2.5×10^6 per 15-cm plate and cultured for 3 days. To induce protein–protein cross-links, cells were first fixed for 30 min on ice with 5 mM dimethyl 3′-dithiobispropionimidate (DTBP, Pierce Biotechnology, Rockford, IL, USA) in cold PBS (pH 8).[44] After two washes in cold PBS (pH 8), remaining DTBP activity was stopped by incubation in DTBP quenching buffer (100 mM Tris–HCl, pH 8, 150 mM NaCl) for 10 min on ice. Next, cells were fixed by addition of 1% formaldehyde in cell culture medium and 10-min incubation at RT. Fixation was stopped by addition of 0.125 M glycine and 5-min incubation at RT. After washing twice in PBS, cells from four to six plates were harvested with a cell scraper in 10 ml SDS buffer (0.1 M NaCl, 50 mM Tris–HCl pH 8, 5 mM EDTA pH 8, 0.2% NaN_3, 0.5% SDS) containing Pefabloc and 1 × Complete Mini protease inhibitor cocktail (Roche). Cells were pelleted at 1200 r.p.m. and resuspended in 2 ml IP buffer (1 volume SDS buffer:0.5 volume Triton dilution buffer (0.1 M Tris–HCl pH 8.6, 0.1 M NaCl, 5 mM EDTA pH 8, 0.2% NaN_3, 5% Triton-X-100)). Cells were

sonicated for 6×30 s with a Branson Sonifier (Sonifier, Danbury, CT, USA) to obtain DNA fragments of 500–1000 bp. The lysates were precleared for 2 h with 4 Fast Flow protein A beads (GE Healthcare, Princeton, NJ, USA) that were blocked o/n (0.2 mg/ml salmon sperm DNA, 0.5 mg/ml lipid-free BSA) and washed in IP buffer. After removal of beads, lysates were diluted in IP buffer corresponding to $\sim 5 \times 10^6$ cells/ml and 1-ml aliquots were incubated rotating o/n at 4°C with primary antibody against ELK1 (Santa Cruz sc-355 X) or rabbit IgG (Sigma I 8140). In addition, 10 μl aliquots were saved as total control samples (1%). Immune complexes were recovered by incubation with blocked protein A beads for 4 h at 4°C and precipitated at $1800 \times g$ at 4°C. Beads were then washed in cold buffers as follows: thrice in 1 ml mixed micelle wash buffer (150 mM NaCl, 10 mM Tris–HCl pH 8, 5 mM EDTA pH 8, 5% w/v sucrose, 0.02% NaN_3, 1% Triton-X-100, 0.2% SDS), twice in 1 ml buffer 500 (50 mM HEPES, pH 7.5, 1 mM EDTA, 500 mM NaCl, 0.2% NaN_3, 1% (v/v) Triton X-100, 0.1% (w/v) deoxycholic acid), twice in 1 ml LiCl/detergent solution (10 mM Tris–HCl pH 8, 1 mM EDTA, 250 mM LiCl, 0.5% (w/v) deoxycholic acid (sodium salt), 0.5% (v/v) IGEPAL, 0.2% NaN_3) and once in 1 ml TE buffer. Samples were inverted 10 times during each wash. Immune complexes were eluted from beads by o/n shaking incubation at 65°C in 1% SDS, 0.1 M $NaHCO_3$ (both IP and input samples) and protein was removed by proteinase K treatment of the supernatant. DNA was extracted with 1 volume phenol:chloroform:isoamylic alcohol (25 : 24 : 1) and ethanol precipitated. qPCR was carried out with SybrGreen qPCR master mix (Applied Biosystems) using the following primers:

miR-34a locus 1 (L1): *34a ChIP F1* 5′-TGGCACGAGCAGGAAGGAGG-3′ and *34a ChIP R1* 5′-GCAGGACTCCCGCAAAATCTCC-3′.
miR-34a locus 2 (L2): *34a ChIP F2* 5′-AATTGTGTAGCCTCCGTAAGGGGA-3′ and
34a ChIP R2 5′-GAAAGAACTAGCCGAGCAAAACCC-3′.
miR-34a 3′ locus: *34a ChIP 3′ F* 5′-GGACTTCGGAAGCTCTTCTGCG-3′ and *34a ChIP 3′ R* 5′-CACCAAGCCCCTGTGCCTTTT-3′.

miRNA pull-out assay. TIG3 TERT/B-RAF:ER cells were treated for 48 h with 500 nM 4-OHT or equal volumes of EtOH carrier. Cells were then transfected with 30 nM miR-34a duplex, 3′ biotin-tagged miR-34a duplex or a mix of two 3′ biotin-tagged miR-34a duplexes containing a UV-reactive 4-tU nucleotide at positions 7 and 11, respectively. At 24 h after transfection, the cells were irradiated with long UV light (365 nm) for 5 min to induce cross-linking of 4-tU nucleotides to RNA, and immediately after this, total RNA was extracted by TRIzol (Invitrogen). After DNase treatment, 10 μg RNA was incubated for 2 h with streptavidin-conjugated beads (GE Healthcare) and washed with DEPC water. RNA was then purified by TRIzol and analysed using qPCR. To measure enrichment of *MYC* transcript, a standard curve was prepared, and HPRT enrichment was used as a negative control. For quantification, 500 ng RNA was reverse transcribed using TaqMan Reverse Transcription Reagents (Applied Biosystems) according to the manufacturer's instructions. qPCR was carried out using SybrGreen PCR mastermix (Applied Biosystems) and the following primers:

HPRT FW: 5′-AGCCAGACTTTGTTGGATTTG-3′
HPRT REV: 5′-TTTACTGGCGATGTCAATAAG-3′
MYC FW: 5′-CCTCGGATTCTCTGCTCTC-3′
MYC REV: 5′-TTCTTGTTCCTCCTCAGAGTC-3′.

Microarray analyses. For miRNA microarrays, TIG3 TERT/B-RAF:ER cells were treated with 500 nM 4-OHT or equal volumes of ethanol on the day after seeding at 1×10^6 per 10-cm plate, in four biological replicates. Total RNA was harvested 72 h later using TRIzol reagent. miRNA microarray analysis (NCode Multi-Species miRNA Microarray V2, Invitrogen) was carried out at the Microarray Center, Rigshospitalet, Copenhagen University Hospital. Briefly, small RNAs (<200 nucleotides) were isolated from total RNA with the PureLink miRNA Isolation kit (Invitrogen). Purified small RNA concentration was measured with a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA), and 1200 ng of purified small RNA was labelled using the NCode miRNA Labelling System (Invitrogen). Briefly, small RNAs were poly(A)-tailed and ligated to a capture sequence. The tagged and tailed miRNAs were subsequently hybridised to the array following the manufactures instructions. Bound miRNAs were detected by another round of hybridisation of branched DNA structures containing Alexa Fluor 3 or Alexa Fluor 5 dye molecules (Invitrogen). All hybridisations were carried out in a MAUI Hybridization System (BioMicro Systems, Salt Lake City, UT, USA). The arrays were scanned in Agilent DNA Microarray Scanner (Agilent Technologies, Santa Clara, CA, USA) and image files were analysed with Genepixe Pro 6.0 software (Molecular Devices, Sunnyvale, CA, USA) (E-MEXP-2038).

For Affymetrix microarrays, TIG3 TERT/B-RAF:ER cells were treated with 500 nM 4-OHT or equal volumes of ethanol on the day after seeding at 1E6 per 10-cm plate, in three biological replicates. After 2 days of 4-OHT treatment, cells were transfected with LNA-miR-34a or a control LNA-scramble using Lipofectamine 2000 (Invitrogen). Total RNA was harvested 24 h after transfection (3 days of 4-OHT treatment in total) using TRIzol reagent. Affymetrix microarray analysis (HG-U133 Plus 2.0 human) was carried out at the Microarray Center, Rigshospitalet, Copenhagen University Hospital. Briefly, double-stranded cDNA was synthesised from 2 μg total RNA using Superscript Choice System (Invitrogen) with an oligo-(dT) primer containing a T7 RNA polymerase promoter. Biotin-labelled antisense cRNA was subsequently *in vitro* transcribed from cDNA (BioArray High Yield RNA Transcript Labelling kit; Enzo Diagnostics, Farmingdale, NY, USA). After fragmentation at 94°C for 35 min (40 mM Tris, 30 mM MgOAc, 10 mM KOAc), samples were hybridised to arrays for 16 h. Upon washing and staining with phycoerytrin-conjugated streptavidin, the arrays were scanned in an Affymetrix GeneArray 2500 scanner as described in the Affymetrix GeneChip protocol. Data have been deposited at ArrayExpress (E-MEXP-2241).

Bioinformatics analyses. The data from miRNA microarrays were analysed with the BRB-ArrayTools V3.7.0 developed by Dr. Richard Simon and BRB-ArrayTools Development Team (http://linus.nci.nih.gov/BRB-ArrayTools.html). Triplicate probe sets were averaged and global normalisation was used to derive the median of the centre of the log-ratios on each array to adjust for the differences in labelling intensities of the Alexa Fluor 3 and Alexa Fluor 5 dyes. We identified miRNAs that were differentially expressed among the two classes using a random-variance *t*-test. Genes were considered statistically significant if their *P*-value was $\leqslant 0.001$.

The data from the Affymetrix array analyses were filtered to contain only probe sets present in at least one sample, in at least two out of three biological replicates. Replicate probe set intensities were averaged and the data were log2 transformed. Probe set annotation from Affymetrix (HG-U133_Plus_2.na25.annot.csv) was mapped to RefSeq IDs. Probe sets with ambiguous mapping (matching more that a single NM) were discarded and probe sets representing the same NM were averaged, resulting in 10184 RefSeq genes. For clustering analysis, we computed the effect of miR-34a inhibition in regular and senescent TIG3 cells, and selected the top 5% most upregulated and 5% most downregulated genes in each pair of samples (the 4-OHT/LNA-miR-34 *versus* 4-OHT/LNA-scr samples, and the EtOH/LNA-miR-34a *versus* EtOH/LNA-scr samples), and unified the four lists, resulting in 1730 genes (this set was $<20\%$ of the genes because the four lists had some overlap). Hierarchical clustering was performed using MATLAB, resulting in 15 clusters. List of differential genes and their assignment to clusters are in Supplementary Table S3. Functional enrichment analysis was carried out using the DAVID website.[25] Promoter analysis was carried out using AMADEUS.[27]

Acknowledgements. We thank Dr. Robert A Hipskind for kindly providing us with the ELK1 expression construct and Dr. Martin Bushell for kindly providing us with the *MYC* 3′UTR luciferase constructs, pLSV-M3′ and pLSV-M3′MUT. Work in the authors' laboratories has been supported by the EC FP7 funding (ONCOMIRS, grant agreement number 201102. The commission is not liable for any use that may be made of the information herein), the Vilhelm Pedersen and Hustrus Foundation, the Danish National Research Foundation, the Danish Medical Research Council, the Danish Cancer Society and the Association for International Cancer Research. This publication reflects only the authors' views.

1. d'Adda di Fagagna F, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, Von Zglinicki T *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* 2003; **426**: 194–198.
2. Herbig U, Jobling WA, Chen BP, Chen DJ, Sedivy JM. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol Cell* 2004; **14**: 501–513.
3. Takai H, Smogorzewska A, de Lange T. DNA damage foci at dysfunctional telomeres. *Curr Biol* 2003; **13**: 1549–1556.
4. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* 1997; **88**: 593–602.
5. Palmero I, Pantoja C, Serrano M. p19ARF links the tumour suppressor p53 to Ras. *Nature* 1998; **395**: 125–126.

6. Courtois-Cox S, Genther Williams SM, Reczek EE, Johnson BW, McGillicuddy LT, Johannessen CM *et al.* A negative feedback signaling network underlies oncogene-induced senescence. *Cancer Cell* 2006; **10**: 459–472.
7. Di Micco R, Fumagalli M, d'Adda di Fagagna F. Breaking news: high-speed race ends in arrest – how oncogenes induce senescence. *Trends Cell Biol* 2007; **17**: 529–536.
8. Courtois-Cox S, Jones SL, Cichowski K. Many roads lead to oncogene-induced senescence. *Oncogene* 2008; **27**: 2801–2809.
9. Mooi WJ, Peeper DS. Oncogene-induced cell senescence – halting on the road to cancer. *N Engl J Med* 2006; **355**: 1037–1046.
10. Olsen CL, Gardie B, Yaswen P, Stampfer MR. Raf-1-induced growth arrest in human mammary epithelial cells is p16-independent and is overcome in immortal cells during conversion. *Oncogene* 2002; **21**: 6328–6339.
11. Bushati N, Cohen SM. MicroRNA functions. *Annu Rev Cell Dev Biol* 2007; **23**: 175–205.
12. Cho WC. OncomiRs: the discovery and progress of microRNAs in cancers. *Mol Cancer* 2007; **6**: 60.
13. Bommer GT, Gerin I, Feng Y, Kaczorowski AJ, Kuick R, Love RE *et al.* p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr Biol* 2007; **17**: 1298–1307.
14. Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH *et al.* Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell* 2007; **26**: 745–752.
15. Corney DC, Flesken-Nikitin A, Godwin AK, Wang W, Nikitin AY. MicroRNA-34b and MicroRNA-34c are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res* 2007; **67**: 8433–8438.
16. Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N *et al.* Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol Cell* 2007; **26**: 731–743.
17. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* 2007; **447**: 1130–1134.
18. Tarasov V, Jung P, Verdoodt B, Lodygin D, Epanchintsev A, Menssen A *et al.* Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle* 2007; **6**: 1586–1593.
19. Pritchard CA, Samuels ML, Bosch E, McMahon M. Conditionally oncogenic forms of the A-Raf and B-Raf protein kinases display different biological and biochemical properties in NIH 3T3 cells. *Mol Cell Biol* 1995; **15**: 6430–6442.
20. Zhu J, Woods D, McMahon M, Bishop JM. Senescence of human fibroblasts induced by oncogenic Raf. *Genes Dev* 1998; **12**: 2997–3007.
21. Tazawa H, Tsuchiya N, Izumiya M, Nakagama H. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc Natl Acad Sci USA* 2007; **104**: 15472–15477.
22. Shaulian E, Zauberman A, Ginsberg D, Oren M. Identification of a minimal transforming domain of p53: negative dominance through abrogation of sequence-specific DNA binding. *Mol Cell Biol* 1992; **12**: 5581–5592.
23. Chang F, Steelman LS, Lee JT, Shelton JG, Navolanic PM, Blalock WL *et al.* Signal transduction mediated by the Ras/Raf/MEK/ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention. *Leukemia* 2003; **17**: 1263–1293.
24. Loots GG, Ovcharenko I, Pachter L, Dubchak I, Rubin EM. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* 2002; **12**: 832–839.
25. Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; **4**: P3.
26. Sun F, Fu H, Liu Q, Tie Y, Zhu J, Xing R *et al.* Downregulation of CCND1 and CDK6 by miR-34a induces cell cycle arrest. *FEBS Lett* 2008; **582**: 1564–1568.
27. Linhart C, Halperin Y, Shamir R. Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res* 2008; **18**: 1180–1189.
28. Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5′UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 2008; **30**: 460–471.
29. Welch C, Chen Y, Stallings RL. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene* 2007; **26**: 5017–5022.
30. Bagchi A, Mills AA. The quest for the 1p36 tumor suppressor. *Cancer Res* 2008; **68**: 2551–2556.
31. Lodygin D, Tarasov V, Epanchintsev A, Berking C, Knyazeva T, Korner H *et al.* Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle* 2008; **7**: 2591–2600.
32. Ohtani N, Zebedee Z, Huot TJ, Stinson JA, Sugimoto M, Ohashi Y *et al.* Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. *Nature* 2001; **409**: 1067–1070.
33. Kumamoto K, Spillare EA, Fujita K, Horikawa I, Yamashita T, Appella E *et al.* Nutlin-3a activates p53 to both down-regulate inhibitor of growth 2 and up-regulate mir-34a, mir-34b, and mir-34c expression, and induce senescence. *Cancer Res* 2008; **68**: 3193–3203.
34. Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, Lund AH. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J Biol Chem* 2008; **283**: 1026–1033.
35. Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* 2005; **438**: 685–689.
36. Raver-Shapira N, Oren M. Tiny actors, great roles: microRNAs in p53's service. *Cell Cycle* 2007; **6**: 2656–2661.
37. Wei JS, Song YK, Durinck S, Chen QR, Cheuk AT, Tsang P *et al.* The MYCN oncogene is a direct target of miR-34a. *Oncogene* 2008; **27**: 5204–5213.
38. Leucci E, Cocco M, Onnis A, De Falco G, van Cleef P, Bellan C *et al.* MYC translocation-negative classical Burkitt lymphoma cases: an alternative pathogenetic mechanism involving miRNA deregulation. *J Pathol* 2008; **216**: 440–450.
39. Kong YW, Cannell IG, de Moor CH, Hill K, Garside PG, Hamilton TL *et al.* The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc Natl Acad Sci USA* 2008; **105**: 8866–8871.
40. Zhuang D, Mannava S, Grachtchouk V, Tang WH, Patil S, Wawrzyniak JA *et al.* C-MYC overexpression is required for continuous suppression of oncogene-induced senescence in melanoma cells. *Oncogene* 2008; **27**: 6623–6634.
41. Wu CH, van Riggelen J, Yetil A, Fan AC, Bachireddy P, Felsher DW. Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. *Proc Natl Acad Sci USA* 2007; **104**: 13028–13033.
42. Guney I, Wu S, Sedivy JM. Reduced c-Myc signaling triggers telomere-independent senescence by regulating Bmi-1 and p16(INK4a). *Proc Natl Acad Sci USA* 2006; **103**: 3645–3650.
43. Littlewood TD, Hancock DC, Danielian PS, Parker MG, Evan GI. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res* 1995; **23**: 1686–1690.
44. Fujita N, Wade PA. Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. *Methods* 2004; **33**: 81–85.

Supplementary Information accompanies the paper on Cell Death and Differentiation website (http://www.nature.com/cdd)

Likely evolutionary origins of microRNAs in animals

Dvir Dahary*#, Reut Shalgi* and Yitzhak Pilpel#

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.

* These authors equally contributed to this work

To whom correspondence should be addressed. Email: dvir.dahary@gmail.com; Pilpel@weizmann.ac.il

Abstract

miRNAs are considered nowadays to be major contributors to the evolution of animal morphological complexity. Studies revealed several episodes of expansion in the number of miRNA families, which correlate with the evolutionary branching of vertebrates, mammals and primates. However, despite the enormous potential role ascribed to miRNAs in evolving animals complexity, their origin, i.e. the evolutionary mechanism that gave rise to them, is not yet fully understood. In this study, we outline two alternative genomic sources which serve as a constant supply for novel miRNAs during evolution. The first is Transposable elements (TEs), which were previously described in the literature as a class of genomic elements serving as an origin for miRNA innovations. The second source is a newly proposed possible origin for miRNAs in evolution: CpG islands (CGIs). We show that these two origins have opposite dynamics during evolution, the TE serving as a source for many novel miRNAs which also rapidly decay, whereas the CGIs give rise to miRNAs which are preferentially retained during evolution. Together our results shed light on the intriguing origin of one of the major constituents of regulatory networks in animals, microRNAs.

Introduction

The dramatic increase in morphological complexity in animal evolution has recently been attributed to non-coding RNAs (ncRNAs) and specifically microRNAs, postulating that they represent a principal layer of gene regulatory networks in metazoans [1-3]. microRNAs (miRNAs) are short endogenous ncRNAs that post-transcriptionally regulate gene expression through interaction with their mRNA targets. Hundreds of animal miRNAs have been identified in the past decade, potentially affecting thousands of mRNA targets, and playing key roles in various pivotal developmental and cellular processes. Recent studies revealed several episodes of expansion in the number of miRNA families, which correspond to the introduction of novel morphological features during animal evolution [1-3]. Remarkably though, despite the enormous potential role ascribed to miRNAs in evolving animals complexity, their origins, i.e. the evolutionary mechanism that gave rise to them, is not fully understood.

Generally, the miRNA pathway is one of three branches of the RNA interference (RNAi) system together with the small-interfering (si)RNA and piwi-interacting (pi)RNA pathways. The RNAi pathway is present in almost all eukaryotes, hence it was suggested that the ancestral RNAi system primarily functioned in defense against viruses and transposable elements (TEs) [4]. In fact, both siRNAs and piRNAs were shown to have key roles in transposon defense [5]. Recently, a few studies suggested the TE-associated origin of several mammal- and primate-specific miRNAs, and further implicated them in TE defense. Smalheiser and Torvik [6] reported 11 cases of presumably TE-derived mammalian miRNAs that showed sequence complementarity with many mRNAs that harbor copies of TEs in their 3' UTRs. Some of these pre-miRNAs hairpins were formed by the junction of two adjacent similar TE segments

residing in opposite orientation. This is one plausible model for miRNA formation directly from the TE sequence. A later work by Piriyapongsa et al. [7] identified dozens of miRNAs overlapping TEs in the human genome, comprising ~12% of the human miRNAs which were then documented. These miRNAs overlapped TE sequences from all four major classes including SINEs, LINEs, LTRs and DNA transposons, suggesting that the formation of novel miRNAs from these elements have occurred in several events during the human genome evolution. Their findings imply on the role of TE-derived miRNAs in the defense against the activity of TEs in the genome. Further supporting this model, Smalheiser and Torvik [8] have reported on 3' UTR-residing *Alu*s as targets of TE-derived miRNAs. Moreover, Lehnert et al. [9] examined the relationship between miRNAs and *Alu* sequences in the human genome and found evidence for the role of miRNAs in guarding against the deleterious effects of *Alu* in the genome. They show that young miRNAs target mainly sense *Alu* (showed significant sequence complementarity against sense *Alu* transcripts as opposed to antisense *Alus*) and are expressed specifically in placenta and testis, where *Alu* may be active. Older *Alu*-associated miRNAs, in contrast, were more widely expressed, suggesting their roles in other processes.

The above reports and others might explain the birth of some novel miRNA families along animal evolution. Here we use the evolutionary classification of miRNA families and set to explore other mechanisms through which novel miRNA families were introduced in animal genomes. We suggest two different origins for novel miRNAs during evolution: the first is the well documented TEs, and the second is CpG rich regions and specifically CpG Islands (CGIs). We also show that these two sources vary differently in their dynamics of miRNA retention during evolution, thus

shedding light on the evolutionary processes that created the human miRNA collections observed today, and suggesting it is still very much under selection.

RESULTS and Discussion

Human miRNA dataset and lineage classification

As a basis for the analyses described below, we classified 670 human miRNAs from miRbase 13.0 by means of three distinct approaches. First, individual miRNAs with a likely common ancestor were grouped into miRNA families. Independently, miRNAs that reside close to each other in the human genome were grouped into miRNA clusters (see Methods). We then further classified both families and clusters into four distinct groups by their probable evolutionary age as inferred by their identification in various genomes throughout the animal kingdom. We designated primate-, mammal- and vertebrate-specific and older miRNAs (the ones that originated before vertebrates radiation, see Methods). We determined the age of both miRNA families and miRNA clusters according to the oldest miRNA sets the class. Our final dataset is described in Table 1.

TE-associated miRNAs

Examining the genomic locations of miRNAs in the human genome, we first set to identify the ones that overlap TE-derived genomic repeats and hence are likely to have originated from ones. All in all, 147 miRNAs were found to overlap genomic repeats from the four main classes - SINEs, LINEs, LTRs and DNA transposons - as mapped in the UCSC genome browser [10] applying RepeatMasker [11] (see Tables S1, S2). These comprise of 22% of the current collection of documented human miRNAs, and is much more than the 12% previously reported [7]. We further

regarded each miRNA family as TE-derived if at least one representative of the family overlapped genomic repeat. We found that most of the TE-associated miRNAs are by and large primate- and mammal-specific. In fact, 31% of the primate-specific and 21% of mammal-specific miRNA families showed association with TEs. In accordance with the assumption that the sequence relics of the TE origin have decayed during the genome evolution, only four vertebrate miRNA families were associated with TEs, and none of the old families. Further supporting the hypothesis that these TEs are the origin of these relatively young miRNA families, is the concurrence between the age of the miRNAs and the presumed age of the TE they overlapped. For instance, primate miRNA families were the only ones associated with *Alu* repeats which are primate-specific, while many mammalian miRNA families seem to have originated from the L2 repeat [6] which was highly active during the mammalian radiation [12] (Supp. Fig. 1). Analysis of the type of TE that miRNAs tend to overlap, revealed two dominant classes: MIR (Mammalian Interspersed Repeat) and L2, in agreement with previous reports [7]. These two ancient mammalian repeats overlap 27 and 20 miRNA families respectively, a much higher number than would be expected (6 and 7 respectively for MIR and L2) based on the frequency of these repeats in the human genome (see supplementary table S3). Interestingly, miRNAs are generally considered to be Pol-II transcribed [13], whereas an interesting study has reported a large cluster of *Alu* flanked miRNAs on chromosome 19 to be Pol-III transcribed [14]. It is plausible that miRNAs derived from MIR elements, which are SINEs derived from tRNAs and contain the tRNA promoter [15], may be another subclass of miRNAs which are also transcribed by Pol-III.

Our findings are in agreement with recent reports regarding the association between miRNAs and transposable elements. However, this mode of miRNA innovation elucidates the origin of at most ~30% of the evolutionarily young miRNA families. The ambiguity regarding the origins and evolution of other miRNA families and particularly ancient miRNAs remains.

CpG-associated miRNAs

We now set to examine the hundreds of miRNAs that could not be associated with genomic repeats, searching for other sequence characteristics that might imply on their origin. The basic structural feature of animal miRNAs and specifically their precursors is a long stable hairpin. In the case of TE-derived miRNAs, it is quite obvious how this hairpin was formed, as evident by manual examination of the overlapping repeats. Many of the TE-associated miRNAs clearly suggest that their hairpins were created by the complementarity between two copies of similar repeats that reside close to each other on opposite strands [6]. Most of animal miRNA precursors, however, are characterized by imperfect hairpins [5]. Assuming that these hairpins require a relatively high GC-content for their stability, we set to analyze the nucleotide content of all human miRNA precursors and their flanking genomic regions. Indeed, The average GC-content of human miRNA hairpins is ~50%, which is considerably higher than the average GC-content of the entire genome (41% [16]). Moreover, the GC-content of the TE-associated miRNAs is clearly lower than that of the remaining miRNAs further implying on their distinctive origins.

Recent studies reported the association between clusters of CpG dinucleotides, designated CpG islands (CGIs), and miRNAs in the context of methylation of miRNA

genes in human tumors [17, 18]. Concurrently, we would like to hypothesize here, that some miRNAs were actually originated from these unique CpG-rich regions.

To test our hypothesis, we looked for miRNAs that physically overlap annotated CGIs in the human genome. Strikingly, we found 65 human miRNAs overlapping CGIs, none of them belongs to the group of TE-associated miRNAs (see Tables S1, S2). Thus, more than 12% of the remaining miRNAs reside within these highly CpG-rich regions. Examining these findings in the context of miRNA families and their evolutionary age, the results are even more surprising.

Notably, CGIs occupy less than 1% of the human genome, setting the ~12% overlap between miRNA precursors and CGIs highly significant (P-value = $8*10^{-57}$; see Methods.). Controlling for possible bias as miRNAs are often transcribed as polycistrons and lie in close proximity to each other on the genome, we still observed more than 12% of miRNA clusters are CGI-associated (p-value = $2*10^{-49}$). We show that this over-representation is highly significant even when comparing it to a variety of different background models which take into account other constraints that may occur on miRNA genomic localization, demonstrating that the observed phenomenon is significant by itself and not as a by-product of a third-party genomic feature (see Methods).

We then compared the origins of human miRNAs, after we classified them according to their evolutionary age. Remarkably, while approximately 10% of primate, mammalian and vertebrate miRNA families are CGI-associated, one third of the old, i.e. the most conserved miRNA families that are common to species from across the animal kingdom, are found within CGIs (Figure 1).

Taking into account that the definition of CGI uses somewhat arbitrary thresholds, we turned to examine the frequencies of dinucleotides in the flanking regions of miRNA

clusters (+/-2Kb, masked for TEs and exons) and compared them to the genomic averages [19]. miRNA clusters were used in this case since genomic sequences were collected, and thus we had to avoid sequence redundancy resulting from consideration of proximal miRNAs. In addition, we excluded the sequence of the miRNA itself, reasoning that miRNAs are overall highly GC rich and their sequence is subjected to other selection constraints. Interestingly, only for a specific dinucleotide – CpG we found a very unique pattern of enrichment, as demonstrated in Fig 2A (for all dinucleotide profiles see Sup Figure S2). In a window of ~300bp upstream and downstream the miRNA position, the observed-to-expected ratio of CpG is significantly higher than the average, peaking at ~+/-60 bps from the miRNA position. Intriguingly we observe that the flanking region of the miRNAs is generally TpA poor, but especially in the +/-150 nucleotides around the miRNA position - a pattern of gradual TpA depletion is observed (Figure 2B). Interestingly, it was reported that the pri-miRNA processing enzyme Drosha, which cleaves the pre-miRNA from its long primary transcript, requires approximately one more helical turn in the miRNA stem in order to efficiently cleave the pre-miRNA form [20]. Our data may indicate that not only high GC content, but rather high CpG content and low TpA content, are required for a stable stem structure of a miRNA and perhaps are also favorable by the Drosha enzyme.

Comparing the distribution of CpG observed/expected values in the regions of miRNAs of different lineages, we see that our previous finding is further strengthened. We observed that the distribution of CpGs in old miRNAs is markedly shifted to the higher values compared to the others (Figure 2C). Even when we filtered out all the CGI-associated miRNAs from all lineages, we still observed a distribution with higher CpG observed/expected values in the old lineages, indicating

traces of more additional miRNAs that might have originated from a CGI that was subsequently decayed (Figure 2D).

Model of two parallel routes of miRNA innovation during evolution

Following the findings described above, we would like to propose a model for the likely origins of miRNA innovations along animal evolution. This model includes at least two discrete routes as illustrated in Figure 1. We suggest here that both the activity of transposable elements and the existence of CpG-rich regions are accountable, independently, for the constant supply of novel miRNA families throughout animal evolution. Evidently, these two evolutionary origins are completely separated as there were no individual miRNAs that were associated with both TE and CGI, and only one miRNA family, miR-220, which has two different members with the two different origins: miR-220b which is vertebrate miRNA which is CGI-associated, and a younger member, miR-220c, which is TE-associated. In general there is only a minor overlap between TEs and CGIs in the human genome. In one route, copies of an active TE serve as a template for hairpin formation, and subsequently the processing into a novel miRNA that is naturally complementary and thus target primarily the TE. Previous reports already suggested that the primary function of these TE-derived miRNAs is probably repression of the potentially deleterious transposition [9]. According to our evolutionary model, once the TE is no longer active, such a miRNA could either acquire some new functional roles or decay over time as it is no longer required. In another route, CpG-rich region, once transcribed, have the potential to form a stable hairpin owing to the presence of self-complementary CG dinucleotides. Next, this hairpin is engaged by the miRNA machinery and assumes a cellular role, namely regulating mRNAs of protein coding

genes. In case it is advantageous to the host, the miRNA will be fixated and will probably be conserved in most descendant species in the lineage.

We further examined another important feature of miRNAs, namely their dependent versus independent transcription as evident by their genomic location. miRNAs can be either *intronic*, i.e. reside within an intron of an existing gene and being expressed with it, or *intergenic*, residing outside the boundaries of known genes. In less frequent cases, miRNAs reside within exons or on the opposite strand of a protein coding gene. In an evolutionary perspective, there seem to be a trend for miRNAs to reside more and more inside introns. Where only ~20% of old miRNA clusters are intronic, and more than 60% are intergenic, equal fractions of vertebrate miRNAs are intronic and intergenic, while mammalian and primate miRNAs are mostly intronic (~50%, see Figure 3A). This trend might be attributed to the biased origin of these groups. Intriguingly, while 36% of the total human miRNA collection are intronic, only 22% of the CGI-associated miRNAs reside within known genes comparing to 50% of the TE-associated miRNAs (Figure 3B). These findings imply on a preference for TE-derived miRNAs to exploit the expression of an existing gene for their function and perhaps for their generation. Presumably, the birth of a novel TE-derived miRNA requires that a copy of the TE will reside in an intron of an already transcribed gene. In this scenario, the insertion of the TE is usually neutral, and it can accumulate mutations that will enable it to be engaged by the miRNA machinery. CGI-associated miRNAs, on the other hand, reside in unique regions with the intrinsic potential to be transcribed independently [21].

TE-originated miRNAs are tissue specific while CGI-miRNAs tend to be broadly expressed

We examined the tissue expression of miRNAs from different lineages and different origins, in 17 normal tissues, using a published expression atlas of miRNAs [22]. When we look at the distributions of the number of tissues in which miRNAs from different lineages are expressed, we see that the older the miRNA, the more tissues (out of the 17 tissues in the data) it is expressed in (Figure 4A). This is not surprising since we expect miRNAs which were originated before the divergence of the vertebrate lineage, to be non tissue specific. When we examine the tissue specificity according to the origins of miRNAs (Figure 4B) we see that TE-originated miRNAs tend to be more tissue specific than other miRNAs, and most of them are not expressed at all in any of the 17 normal tissues examined (KS-test p-value = 0.0053). As many of the primate-specific miRNAs were not tested when the expression atlas was published, we had to unify primate and mammalian miRNAs in order to get a more detailed picture on the difference between CGI-originated and TE-originated miRNAs, with an attempt to neutralize the effect of the age of the miRNA as much as possible (Figure 4C). Here too we see a slight, however non significant, tendency for CGI-derived miRNAs to be less tissue specific compared to TE-derived miRNAs. Interestingly, here the miRNAs which were not assigned to either origin have a tissue specificity distribution which is more similar to the TE-derived.

Differential evolutionary dynamics of miRNA innovation from the two origins

We identify two opposite trends in the origins of miRNAs during evolution. TE-originated miRNAs are highly frequent in the primate lineage, less so in the mammalian lineage, and not observed at all in vertebrate or old miRNAs. On the other hand, CGI-originated miRNAs are a large fraction of the old miRNAs. These two opposite trends may suggest that whereas CGI-originated miRNAs tend to be more

conserved and retained in evolution, since their fraction increases with miRNA evolutionary age, the repeat-originated miRNAs are being born at high rates but also decay much faster. It is possible that repeat-originated miRNAs are retained in evolution while the repeats that surround them are actually lost, and therefore some of the vertebrate-specific and older miRNAs were actually originated from repeats. This claim can also be true for the CGI origin; however, in this case we can identify those relics as regions that retained high rates of CpG dinucleotides, as mentioned above. In the case of repeats, we cannot do that in such a straightforward manner. We suggest that this is indeed the case for some of the miRNAs, namely that miRNAs are constantly born from repeats and from CGIs and are selected for and thus retained during evolution. However, the CGI-originated miRNA are more likely to be retained in evolution, whereas repeat-originated miRNAs are rapidly born and also decay at high rates along with the repeats that host them. Alternatively, it is possible that repeat-originated miRNAs are retained in evolution while the repeats that surround them are actually lost, and therefore some of the vertebrate-specific and older miRNAs were actually originated from repeats that cannot be detected anymore. However some evidence point to the other option. Interestingly, a study comparing several *Drosophila* species has reported a similar evolutionary trends for fly miRNAs [23]. Lu et al. have shown that the birth and death rate of novel miRNAs in *Drosophila* is very high, and they make a distinction between two classes of "functional" and "non-functional" miRNAs. They indicate that the novel fly miRNAs emerge from non-miRNA sequences, i.e. not through the expansion of existing miRNA families, however they did not suggest potential origins for novel miRNAs.

To support of our model of rapid death for TE-derived miRNAs, we compared the numbers of repeat-originated miRNAs in the primate and mammalian lineages to their

expected rates of death, according to their originators, the TEs themselves. We classify the miRNA originating TEs to lineages according to [24], which basically indicated that, out of all originating TEs, only *Alu* repeats are primate-specific. We see that 108 primate-specific miRNAs are TE-originates, corresponding to 68 novel miRNA families, whereas only 34 mammalian specific miRNAs are TE-originated, corresponding to 31 miRNA families: more than three times less, or more than twice less when we consider miRNA families (see table S3). However, when we counted the total sequence of all mammalian and primate repeats which ever gave rise to miRNAs, we see that while mammalian repeats occupy ~1Gbps of the human genome, the primate repeats we counted occupy only ~0.3GBps. Given that mammalian repeats are capable of serving as origins for primate specific miRNAs too, the potential of repeat sequence that could theoretically give rise to primate specific miRNAs is the sum of the two, ~1.3Gbps (see table S3). Thus, we see that while mammalian repeat sequences that were retained in the human genome are only 1.3 less than the total TE-sequence in the human genome, or the estimated death rate of TEs between humans and mammals is 1.3, the death rate of TE-originated miRNAs is much higher – 2.19 to 3.17. This may indicate that indeed miRNAs are generated from TEs across evolution, but disappear much more rapidly than expected by the selection forces that are applied on TEs, and support our model. Extrapolating from these numbers, we may guess that the current collection of human miRNAs is still under selection, and given these rates, out of the current 108 TE-derived human specific miRNAs, only ~41-45 will eventually be further retained in evolution.

Taken together, our results support the hypothesis that novel miRNA hairpins originated several times during animal evolution from CpG-rich regions. Moreover

we propose that, as opposed to TEs which have a high death rate, CGI-originated miRNAs are preferentially retained in evolution. This is the first evidence that sheds light on an intriguing newly proposed origin of one of the major constituents of regulatory networks in animals, microRNAs. Surprisingly, these findings also suggest a primary role for emergence of CpG-rich regions in the evolution of animal complexity.

Methods

Sequence and genome annotations

We examined the genomic locations of 676 human miRNAs and first excluded the 4 miRNAs that has multiple genomic positions, and other 2 miRNAs that were excluded from miRNABase. microRNA (miRNA) and their RNA hosts, CpG islands (CGIs) and genomic repeats data were downloaded from the UCSC annotation database for the human genome (hg18; [10]).

Genomic clusters of miRNAs were defined as neighboring miRNAs with <10Kb genomic distance as described in Shalgi et al. [25]. A cluster was considered as TE / CGI-associated if at least one of its miRNA members was TE / CGI-associated.

miRNA families were grouped based on miRNA names, as in [3].

Significance of the association between miRNAs and CGIs

As a first means to evaluate the statistical significance of the association between miRNAs and CGIs we simply computed the probability of genomic segments with the average length of miRNAs to overlap CGIs by computing the Poisson distribution as described below. Taking into consideration 28226 CGIs of total length of 21.5 Mbps,

and euchromatic genome size of 2.85 Gbps, the probability to find at random 65 miRNAs or more that overlap a CGI is $8.2*10^{-57}$.

Next, we wished to further control for any local genomic characteristics which could be associated with CGIs, such as high GC content, expression potential, genomic isochors, and regulatory promoter regions. Such factors might indirectly influence the association we observe.

In addition, we wanted to compare these miRNAs to a genomic background which bears similar characteristics to their natural genomic environment, in order to control for any potential bias towards other CGI related features that might influence the association observed. Furthermore, a potential concern is that the significant global association of miRNAs with CGIs, is influenced by intronic miRNAs, which reside in regions with a background tendency to reside in vicinity of CGI. We thus chose to focus on a set of 258 intronic miRNAs, comprising 38% of all analyzed miRNAs, and compared them to a genomic background set of all the other introns of their host genes (3429 introns). While 10% of all miRNAs are CGI-associated, the fraction in the intronic miRNAs is actually lower, 6%, while 12% of the intergenic miRNAs are CGI-associated. We found that the tendency of intronic miRNAs to be associated with CGIs even on the background of their own genomic context is still highly significant (P-value = $2*10^{-11}$, Poisson statistics, see below). Therefore, we can conclude that the significant miRNA-CGI association observed, is not due to the association of CGI with transcribed regions in general, or specifically with the genomic regions from which human miRNAs are transcribed.

A cumulative Poisson p-value was calculated for the probability of finding the observed, or higher, number of miRNAs residing within CpG islands , based on the fraction of sequence of background intronic sequence, which overlaps CGIs:

$$pv = \sum_{x=X}^{total\ \#ofmiRs} Poisson(x, \lambda)$$

where X equals the number of miRNAs associated with CGIs, and

λ = total # of miRNAs × total length of CGIs in background / total background length

Significance of the association between miRNAs and CGIs – additional controls

We also addressed the concern that the effect we see is due to the possible tendency of miRNAs to reside within the upstream introns, which are more likely to be in the vicinity of promoter-associated CGIs of the host genes. We repeated the analysis when the background for the statistical analysis was derived only from introns upstream to the miRNA-hosting intron (Table S4). With this background too we received a significant p-value ($3.2*10^{-10}$).

Dinucleotide frequencies

Sequences were taken from hg18 version of the genome, 2kb upstream and downstream to each miRNA, excluding the pre-miRNA sequence. For miRNA clusters we took the sequence of the entire cluster, excluding the miRNAs within it, plus 2kb flanking sequences. Then, we masked all repeats from the LINE, SINE, DNA and LTR classes, and masked coding exon sequences when necessary.

Dinucleotide frequency (figures 2 and S2) was calculated as its number of occurrences, divided by the window size, and then normalized to the product of the two individual nucleotide fractions (count divided by window size) in this window.

For profile presentation, these calculations were performed for each dinucleotide (16 in total) for each miRNA. Figure S2 (and figure 2A and 2B) present, for each dinucleotide, over a stretch of 2kb flanking sequence of the miRNAs, in a running window of 100bp, the average of the dinucleotide observed/expected, over all miRNAs examined. The red lines indicate the genome averages for each dinucleotide, as taken from (ref). For the profile plots (Figure 2A and 2B), cluster sequences, which were longer than 2kb and could not be centered around one miRNA, were eliminated.

Tissue expression analysis

Expression data was downloaded from [22], and filtered to contain only 17 human normal tissues: hsa_Cerebellum-adult, hsa_Frontal-cortex-adult, hsa_Midbrain-adult, hsa_Hippocamp-adult, hsa_Liver, hsa_Heart, hsa_Spleen, hsa_Pituitary, hsa_Thyroid hsa_Pancreatic-islets, hsa_USSC, hsa_Ovary, hsa_Testis, hsa_Uterus, hsa_Placenta, hsa_Epididymis, and hsa_Prostata.

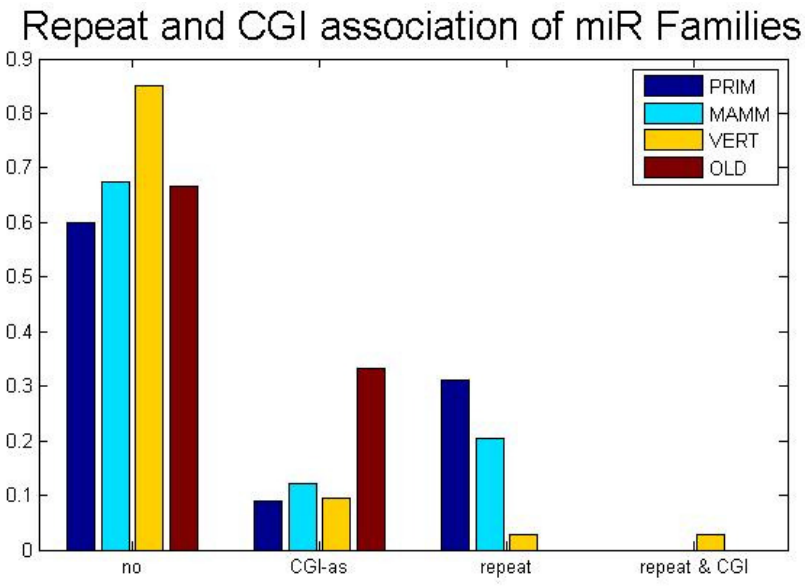
miRNA was considered expressed in a tissue if it had one or more clones in that tissue. miRNAs which did not appear in the data were excluded from the tissue-counts analysis.

	# of miRNAs	**# of miRNA families**	**# of miRNA clusters**
PRIMATE	**284**	**215**	**228**
MAMMALIAN	**175**	**156**	**101**
VERTEBRATE	**107**	**73**	**81**
OLD	**104**	**39**	**69**
Total	**670**	**483**	**479**

Table 1

Summary of the miRNAs, families and clusters that were analyzed in our study.

Figure 1



Fraction of miRNA families from the four lineages which are CGI-associated, TE-associated, none of the two or both. One vertebrate family, miR-220 is associated with both CGI and TE, whereas miR-220b is a CGI associated vertebrate miRNA, miR-220c is a TE-associated mammalian miRNA.

Figure 2A

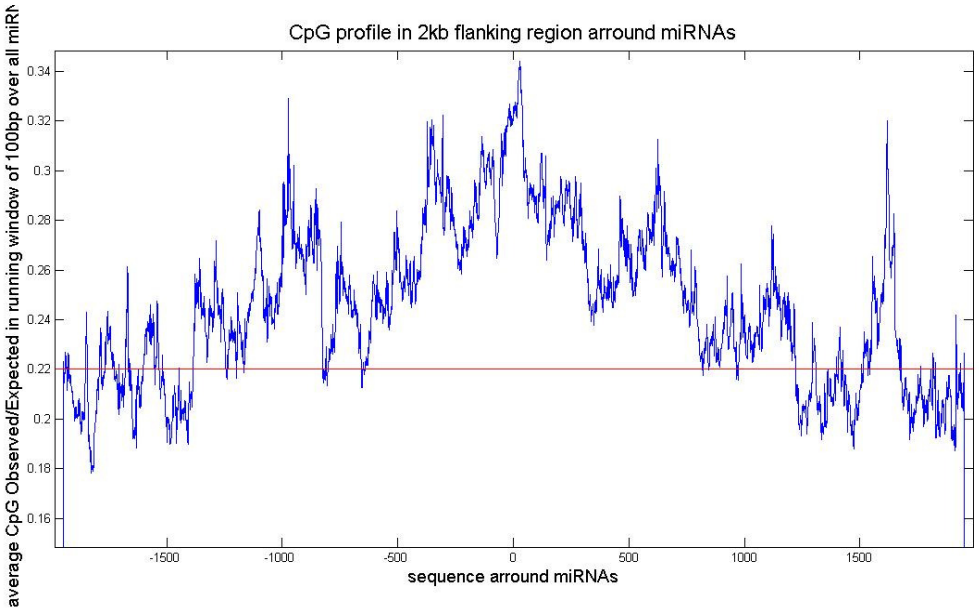


Figure 2B

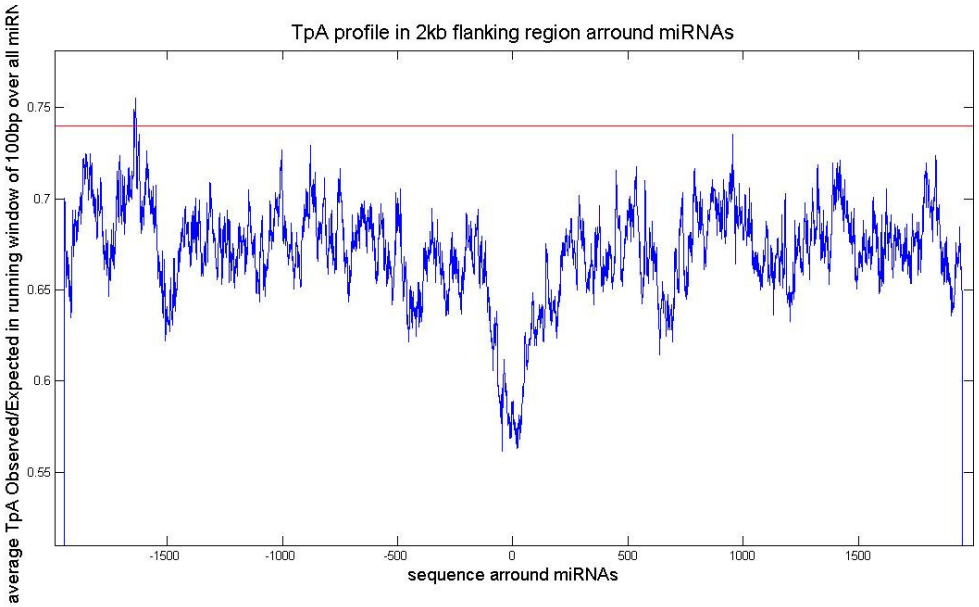


Figure 2C

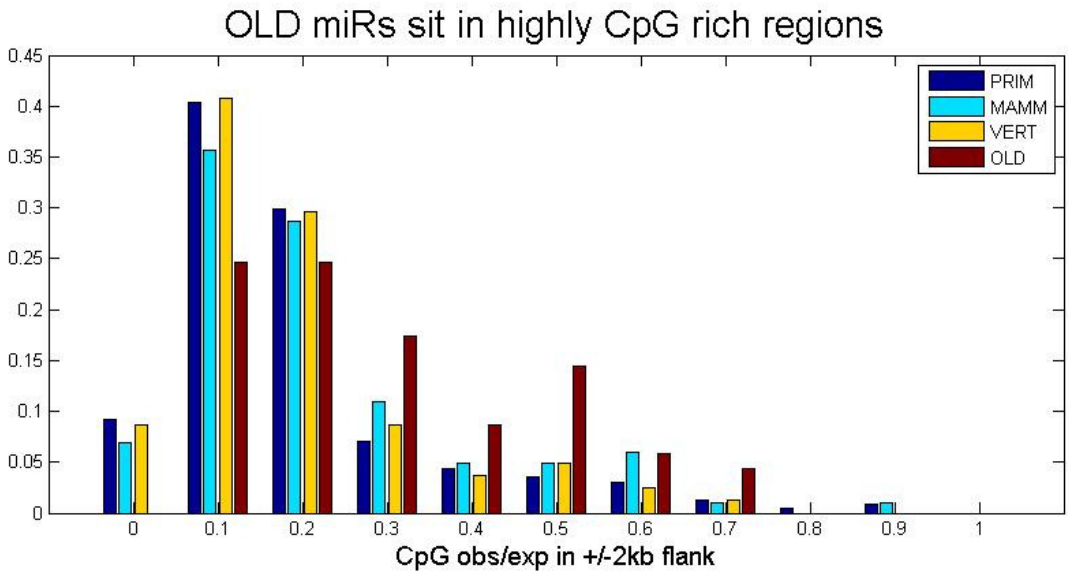
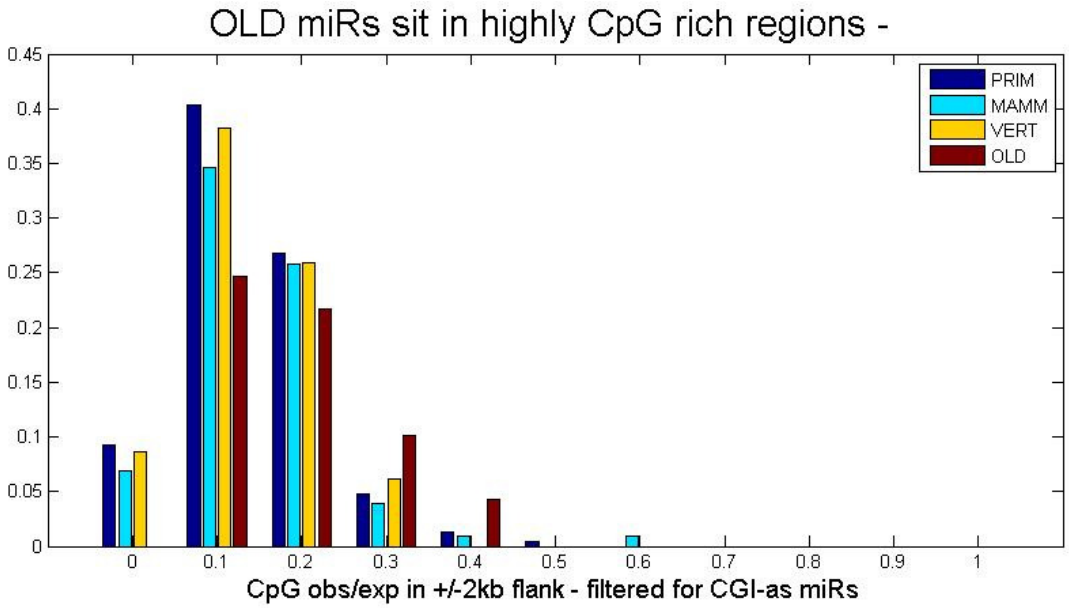


Figure 2D



Average CpG and TpG profile of all individual miRNAs. (**A**) CpG/ (**B**) TpG dinucleotide Observed/Expected was calculated in a running window of 100bp in a total range of +/-2kb flanking each miRNA, excluding the sequence of the miRNA itself. miRNA clusters were discarded from this analysis. Sequences were masked for repeat classes of LINE, SINE, DNA and LTR, and exonic sequences were masked

too. (**C**) Distribution of CpG observed/expected values in miRNA clusters flanking regions. In general, old miRNA clusters have a much higher CpG Observed/expected values. Genome average if 0.22. Fractions are within each lineage.

(**D**) Distribution of CpG observed/expected values in miRNA clusters flanking regions. CGI-associated miRNAs were filtered out of this analysis. Nevertheless, we can still observe the tendency of old miRNA clusters to higher CpG Observed/expected values. Fractions are within each lineage.

Figure 3A

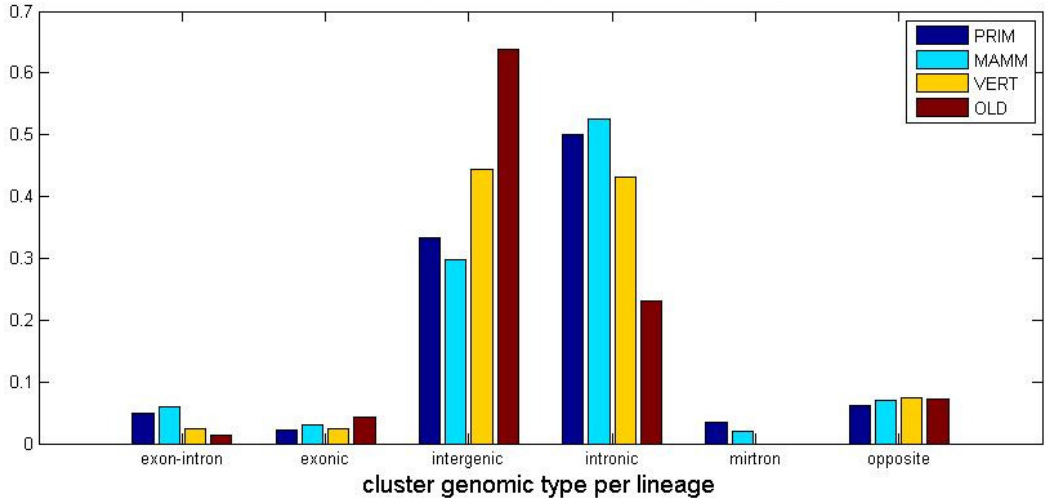
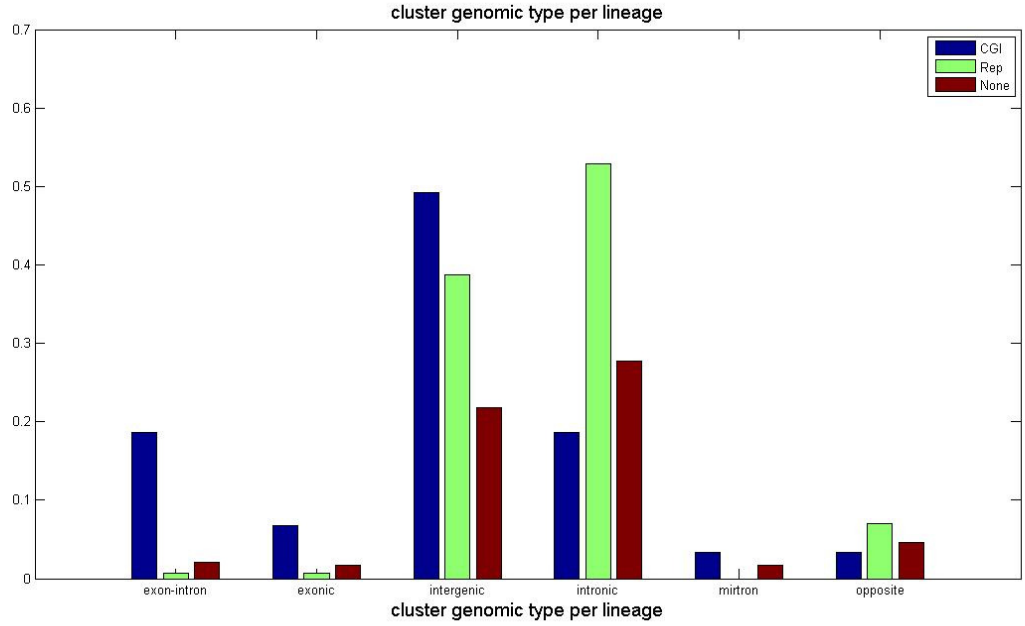


Figure 3B



(**A**) Older miRNAs tend to reside more in intergenic regions while younger ones are more intronic. Classification of miRNA clusters was performed to lineages. Fractions are within each lineage. (**B**) CGI-associated miRNAs are more intergenic while TE-derived miRNAs tend to reside in introns. Fractions are within each origin.

Figure 4A

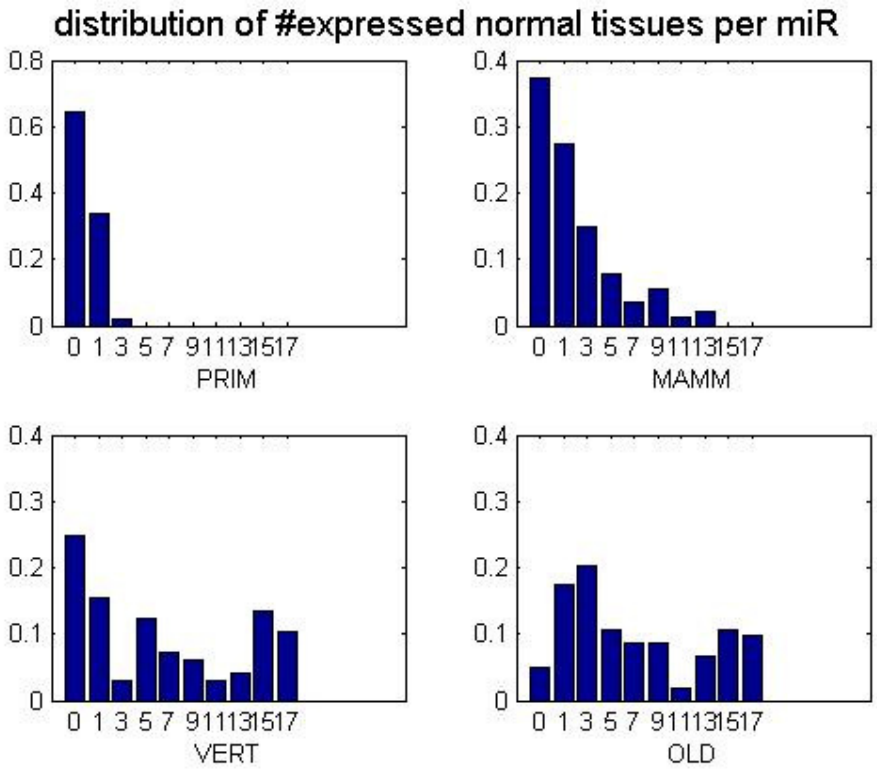


Figure 4B

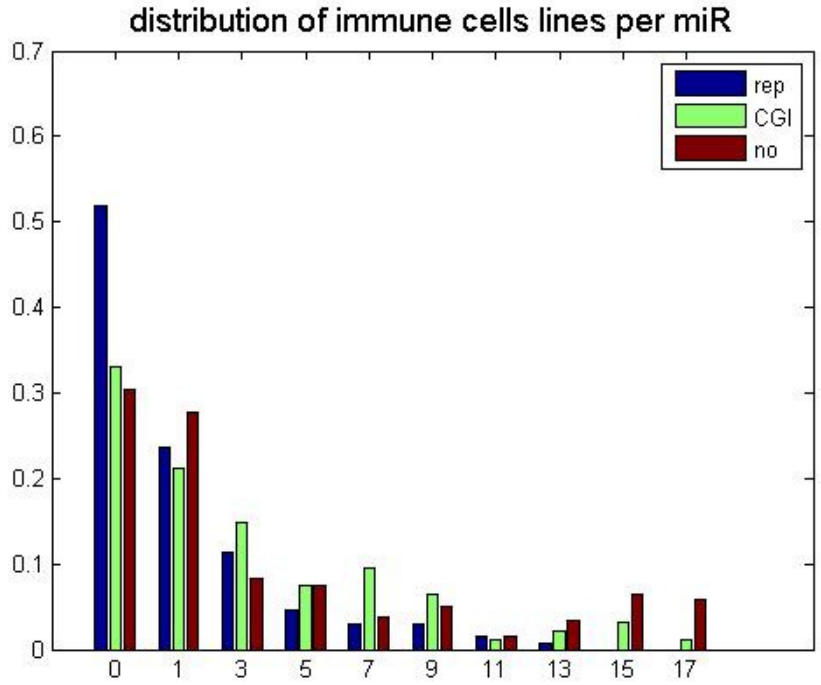
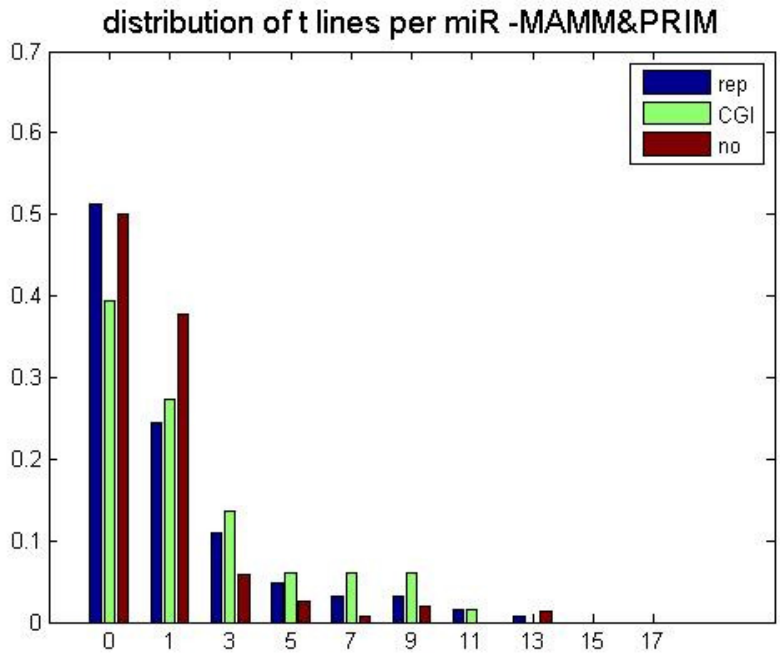


Figure 4C



(**A**) The older miRNAs tend to be more broadly expressed while younger miRNAs are more tissue specific. Distributions of number of tissues each lineage miRNAs are expressed in out of a total of 17 normal human tissues. Fractions are normalized within each lineage. (**B**) The TE-derived miRNAs tend to overall be more tissue specific (or not expressed at all in each of the 17 tissues in the data) than CGI-associated miRNAs. Fractions are normalized to each origin type. (**C**) Same as B but only for mammalian and primate miRNAs.

Figure S1A

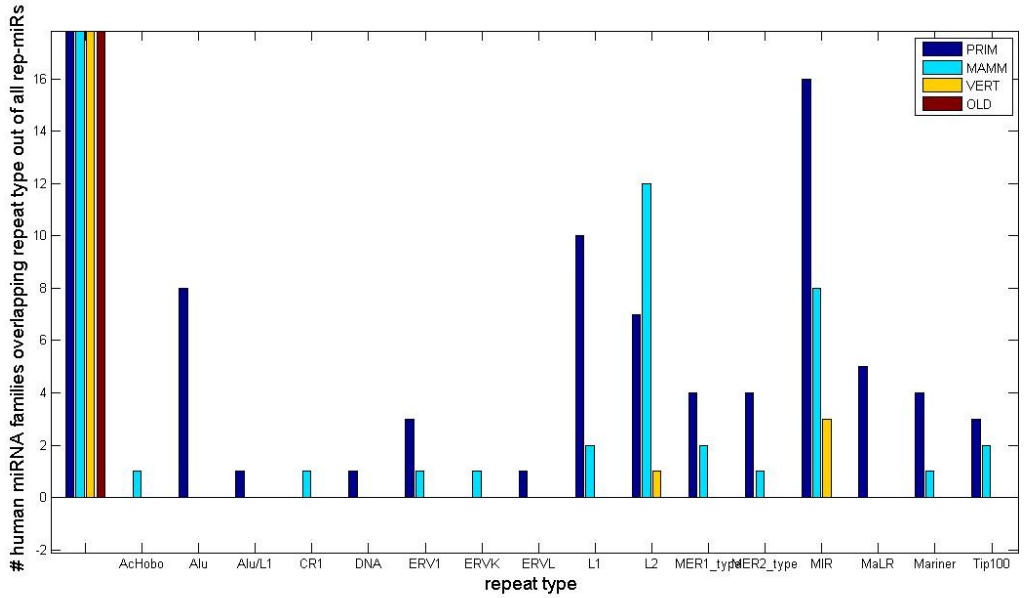
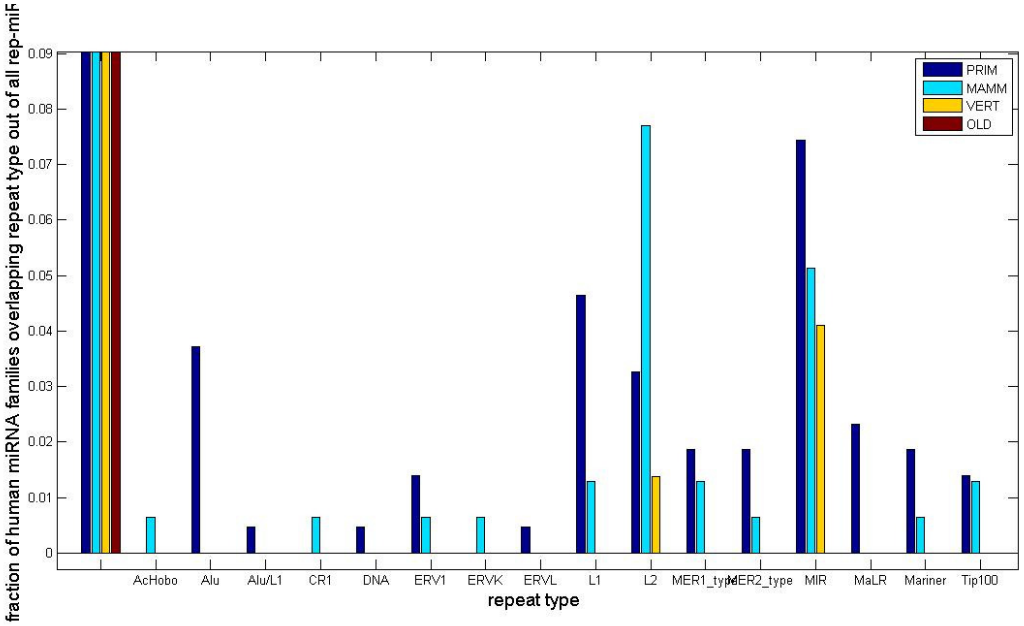
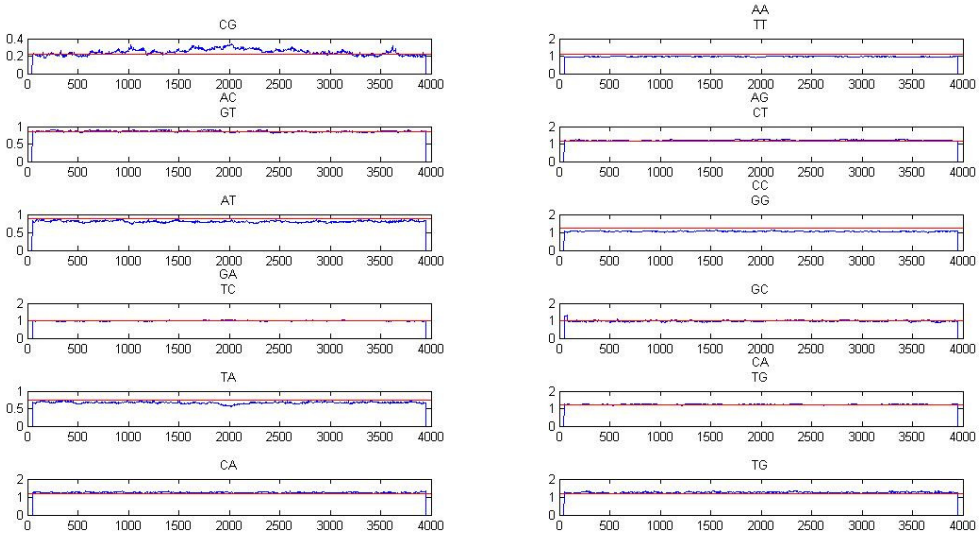


Figure S1B



Numbers (**A**) and fractions (**B**) of miRNAs from the four lineages that overlap different TE families.

Figure S2



Dinucleotide plots were calculated the same as in Figure 2, for all possible 12 dinucleotides.

References

1. Grimson, A, Srivastava, M, Fahey, B, Woodcroft, BJ, Chiang, HR, King, N, Degnan, BM, Rokhsar, DS, and Bartel, DP. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature. 2008; 455: 1193-1197.
2. Niwa, R and Slack, FJ. The evolution of animal microRNA function. Curr Opin Genet Dev. 2007; 17: 145-150.
3. Sempere, LF, Cole, CN, McPeek, MA, and Peterson, KJ. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zoolog B Mol Dev Evol. 2006; 306: 575-588.
4. Cerutti, H and Casas-Mollano, JA. On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet. 2006; 50: 81-99.
5. Shabalina, SA and Koonin, EV. Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol. 2008; 23: 578-587.
6. Smalheiser, NR and Torvik, VI. Mammalian microRNAs derived from genomic repeats. Trends Genet. 2005; 21: 322-326.
7. Piriyapongsa, J, Marino-Ramirez, L, and Jordan, IK. Origin and evolution of human microRNAs from transposable elements. Genetics. 2007; 176: 1323-1337.
8. Smalheiser, NR and Torvik, VI. Alu elements within human mRNAs are probable microRNA targets. Trends Genet. 2006; 22: 532-536.
9. Lehnert, S, Van Loo, P, Thilakarathne, PJ, Marynen, P, Verbeke, G, and Schuit, FC. Evidence for co-evolution between human microRNAs and Alu-repeats. PLoS One. 2009; 4: e4456.
10. Karolchik, D, Baertsch, R, Diekhans, M, Furey, TS, Hinrichs, A, Lu, YT, Roskin, KM, Schwartz, M, Sugnet, CW, Thomas, DJ, Weber, RJ, Haussler, D, and Kent, WJ. The UCSC Genome Browser Database. Nucleic Acids Res. 2003; 31: 51-54.
11. Smit, AF. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev. 1996; 6: 743-748.
12. Ichiyanagi, K and Okada, N. Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. Mol Biol Evol. 2008; 25: 1148-1157.
13. Lee, Y, Kim, M, Han, J, Yeom, KH, Lee, S, Baek, SH, and Kim, VN. MicroRNA genes are transcribed by RNA polymerase II. Embo J. 2004; 23: 4051-4060.
14. Borchert, GM, Lanier, W, and Davidson, BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006; 13: 1097-1101.
15. Smit, AF and Riggs, AD. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res. 1995; 23: 98-102.
16. Lander, ES, Linton, LM, Birren, B, Nusbaum, C, Zody, MC, Baldwin, J, Devon, K, Dewar, K, Doyle, M, FitzHugh, W, Funke, R, Gage, D, Harris, K, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409: 860-921.
17. Lujambio, A, Calin, GA, Villanueva, A, Ropero, S, Sanchez-Cespedes, M, Blanco, D, Montuenga, LM, Rossi, S, Nicoloso, MS, Faller, WJ, Gallagher, WM, Eccles, SA, Croce, CM, et al. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A. 2008; 105: 13556-13561.

18. Weber, B, Stresemann, C, Brueckner, B, and Lyko, F. Methylation of human microRNA genes in normal and neoplastic cells. Cell Cycle. 2007; 6: 1001-1005.
19. Simmen, MW. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. Genomics. 2008; 92: 33-40.
20. Zeng, Y, Yi, R, and Cullen, BR. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. Embo J. 2005; 24: 138-148.
21. Sandelin, A, Carninci, P, Lenhard, B, Ponjavic, J, Hayashizaki, Y, and Hume, DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet. 2007; 8: 424-436.
22. Landgraf, P, Rusu, M, Sheridan, R, Sewer, A, Iovino, N, Aravin, A, Pfeffer, S, Rice, A, Kamphorst, AO, Landthaler, M, Lin, C, Socci, ND, Hermida, L, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007; 129: 1401-1414.
23. Lu, J, Shen, Y, Wu, Q, Kumar, S, He, B, Shi, S, Carthew, RW, Wang, SM, and Wu, CI. The birth and death of microRNA genes in Drosophila. Nat Genet. 2008; 40: 351-355.
24. Lowe, CB, Bejerano, G, and Haussler, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc Natl Acad Sci U S A. 2007; 104: 8005-8010.
25. Shalgi, R, Lieber, D, Oren, M, and Pilpel, Y. Global and Local Architecture of the Mammalian microRNA-Transcription Factor Regulatory Network. PLoS Comput Biol. 2007; 3: e131.

Differentially Regulated Micro-RNAs and Actively Translated Messenger RNA Transcripts by Tumor Suppressor p53 in Colon Cancer

Yaguang Xi,[1] Reut Shalgi,[2] Oystein Fodstad,[1] Yitzhak Pilpel,[2] and Jingfang Ju[1]

Abstract

Purpose: The aim of this study was to investigate the role of p53 in regulating micro-RNA (miRNA) expression due to its function as a transcription factor. In addition, p53 may also affect other cellular mRNA gene expression at the translational level either via its mediated miRNAs or due to its RNA-binding function.

Experimental Design: The possible interaction between p53 and miRNAs in regulating gene expression was investigated using human colon cancer HCT-116 (wt-p53) and HCT-116 (null-p53) cell lines. The effect of p53 on the expression of miRNAs was investigated using miRNA expression array and real-time quantitative reverse transcription-PCR analysis.

Results: Our investigation indicated that the expression levels of a number of miRNAs were affected by wt-p53. Down-regulation of wt-p53 via small interfering RNA abolished the effect of wt-p53 in regulating miRNAs in HCT-116 (wt-p53) cells. Global sequence analysis revealed that over 46% of the 326 miRNA putative promoters contain potential p53-binding sites, suggesting that some of these miRNAs were potentially regulated directly by wt-p53. In addition, the expression levels of steady-state total mRNAs and actively translated mRNA transcripts were quantified by high-density microarray gene expression analysis. The results indicated that nearly 200 cellular mRNA transcripts were regulated at the posttranscriptional level, and sequence analysis revealed that some of these mRNAs may be potential targets of miRNAs, including translation initiation factor eIF-5A, eIF-4A, and protein phosphatase 1.

Conclusion: To the best of our knowledge, this is the first report demonstrating that wt-p53 and miRNAs interact in influencing gene expression and providing insights of how p53 regulates genes at multiple levels via unique mechanisms.

The tumor suppressor gene *p53* is one of the key regulators of cell cycle control and apoptosis and has been named the guardian of the genome (1). In addition to its function as a transcription factor, p53 also acts as an RNA-binding protein capable of regulating its own mRNA translation (2). As an RNA-binding protein, p53 regulates the expression of other cellular mRNA transcripts at the posttranscriptional level (3). p53 also influences apoptosis by accumulating to mitochondria (4, 5).

With the recent discovery of noncoding RNAs [micro-RNAs (miRNA) and small interfering RNAs (siRNA)] and their function as translational regulators, it is clear that miRNAs play important roles in regulating gene expression. The notion that miRNAs regulate gene expression at the translational level is based on the study of the first two miRNAs, *lin*-4 and *let*-7, in *Caenorhabditis elegans*. *Lin*-4 attenuates the translation, but not the mRNA level, of two target genes, *lin*-14 and *lin*-28, by imperfect base pairing to complementary sequences in the 3′ untranslated region of the target mRNAs (6, 7). Translational regulation has been extensively studied in plant biology (8). In plants, translational regulation provides acute responses due to sudden environmental changes and this process is highly reversible and energy efficient. Translational control also provides the same advantage for mammalian systems, in particular during genotoxic stress (9).

The central concept of translational regulation is that gene expression may be controlled by the efficiency of translation of a given mRNA in the absence of a corresponding change in the steady-state level of that mRNA. Translational regulation provides the cell with a more precise, immediate, and energy-efficient way of controlling expression of proteins, and can induce rapid changes in protein synthesis without the need for transcriptional activation and subsequent mRNA processing steps. In addition, translational control also has the advantage of being readily reversible, providing the cell with great flexibility in responding to various cytotoxic stresses.

Authors' Affiliations: [1]University of South Alabama-Cancer Research Institute, Mobile, Alabama and [2]Weizmann Institute of Science, Rehovot, Israel

Received 8/23/05; revised 1/18/06; accepted 2/1/06.

Grant support: University of South Alabama-Cancer Research Institute start-up fund (J. Ju) and Israel Science Foundation (Y. Pilpel).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Jingfang Ju, Cancer Genomics Laboratory, University of South Alabama-Cancer Research Institute, MSB2316, 307 North University Boulevard, Mobile, AL 36688. Phone: 251-460-7393; Fax: 251-460-6994; E-mail: jju@usouthal.edu.

doi:10.1158/1078-0432.CCR-05-1853

Little is known, however, how miRNAs are regulated at the transcriptional level. After transcription, pre-miRNAs are processed by Dicer complex to their corresponding mature miRNAs. We hypothesize that p53 may also mediate certain miRNAs expression due to its function as a transcription factor. In addition, p53 may also affect other cellular mRNA gene expression at the translational level either via its mediated miRNAs or due to its own RNA-binding function. This hypothesis is partially supported by a recent report from O'Donnell et al. (10) showing that c-Myc regulated a number of miRNAs, and two of the miRNAs regulated E2F expression. c-Myc is a helix-loop-helix leucine zipper transcription factor that regulates an estimated 10% to 15% of genes in the human genome.

Translational control has been shown to play a key role in oncogenesis (9). One of the examples is thymidylate synthase, one of the important targets for fluoropyrimidine-based anticancer therapy (11). Another example is vascular endothelial growth factor, which was shown to be regulated, at least in part, at the translational level (12). More importantly, *p53*, the critical tumor suppressor gene, was also regulated at the translational level (2). However, the RNA-binding function of p53 and its potential for regulating other downstream genes has not been fully elucidated.

The main function of miRNAs is to regulate gene expression at the translational level. Although the exact function of most of the newly discovered miRNAs and siRNAs are just emerging, their ability to regulate cell proliferation and cell death has been recently shown (13). Recent reports have shown that expression of miRNAs can be altered in cancer (14). With the recent discovery of the function of miRNA as translational attenuators, we have reasoned that there might be a potential interaction between miRNAs and p53 because of the dual function of p53 as a transcription factor and RNA-binding protein, and the roles of both in the translational regulation process.

Therefore, we chose to explore the potential relationship between the transcription factor function of p53 and miRNA expression in a colon cancer–related context, as *p53* is one of the most frequently altered tumor suppressor genes in colon cancer due to mutations and deletions. The human HCT-116 (wt-p53) and HCT-116 (null-p53) colon cancer cell lines were chosen as model systems to investigate the role of p53 on the expression of miRNAs. HCT-116 (null-p53) cell line was developed via targeted deletion using homologous recombination using HCT-116 (wt-p53) cells (15). This model has been used extensively for the investigation of p53 functions in cell cycle control and apoptosis (15–18). We expect that the functional miRNAs are localized in the actively translated polyribosome complexes (19). Hence, we have investigated the effect of wt-p53 on miRNAs and their translationally regulated mRNA targets by isolating both actively translated mRNA transcripts and miRNAs from polyribosome complexes from these two colon cell lines. The effect of p53 on miRNA expression and on the expression levels of both steady-state and actively translated mRNA transcripts were analyzed. Our study indicated that the expression levels of a number of miRNAs were affected by wt-p53. Down-regulation of wt-p53 via siRNA abolished the effect of wt-p53 in regulating miRNAs in HCT-116 (wt-p53) cells. Global sequence analysis revealed that >46% of the 326 miRNA putative promoters contain potential p53-binding sites, suggesting that some of these miRNAs were potentially regulated directly by wt-p53. Nearly 200 cellular mRNA transcripts were regulated at the posttranscriptional level, and sequence analysis revealed that some of these mRNAs may be potential targets of miRNAs.

Materials and Methods

Cell lines and reagents. The HCT-116 (wt-p53) and HCT-116 (null-p53) cell lines were a gift from Dr. Bert Vogelstein (The Johns Hopkins University, Baltimore, MD) and were described in detail previously (15, 16). Both cell lines were maintained in McCoy's medium supplemented with 10% fetal bovine serum, 1 mmol/L sodium pyruvate, 2 mmol/L L-Glutamine, and antibiotics. All cell lines were grown at 37°C in a humidified incubator with 5% CO_2. 5-Fluorouracil (5-FU) was purchased from Sigma (St. Louis, MO).

Isolation of steady-state total mRNA and actively translated mRNA transcripts. The procedures for isolating steady-state total mRNA and actively translated mRNA transcripts were described in detail previously via sucrose gradient ultracentrifugation (20). The activated translated mRNA transcripts were isolated from pooled polysome fractions (fractions 7-13) using Trizol-LS Reagent (Invitrogen, Carlsbad, CA).

mRNA transcript expression analysis using microarray. CodeLink UniSet Human 20 K Bioarray (GE Healthcare/Amersham Biosciences, Piscataway, NJ), containing ~20,289 gene probes, was used to generate gene expression profiles of both steady-state total mRNAs and actively translated mRNAs isolated from HCT-116 (wt-p53) and HCT116 (null-p53). All reagents and protocols were provided by GE Healthcare/Amersham Biosciences. Double-stranded cDNAs were generated using 2 μg RNA from each sample. After purification, the double-stranded cDNAs were used as templates to generate cRNA via an *in vitro* transcription reaction using T7 RNA polymerase and biotin-11-UTP (Perkin-Elmer, Boston, MA). Biotin-labeled cRNA (10 μg) was fragmented and hybridized to a UniSet Human 20 K Bioarray. The arrays were washed and stained with Cy5-streptavadin. After washing, the dried slides were scanned by Axon GenePix Professional 4200A microarray scanner using Genepix Pro 5.1 software. The images were grided by Codelink 4.1 software (GE-Healthcare/Amersham Biosciences). Contaminated and irregularly shaped spots were removed before the data files were analyzed. GeneSpring Software 7.2 (Agilent, Palo Alto, CA) was used for the final gene expression analysis. Under Cross-Gene Error Model, normalization step was done in two steps: (*a*) "per chip normalization," in which each measurement was divided by the 50th percentile of all measurements in its array, and (*b*) "per gene normalization," in which all the samples were normalized against the specific samples (controls). The results were filtered by flags and 4-fold cutoff. The expression profiles were compared using one-way ANOVA analysis with $P < 0.05$.

Mature miRNA expression analysis using miRNA array. The cDNA synthesis procedures for miRNA quantitation using total RNAs was based on method described by Elbashir et al. (21). Actively translated RNAs from HCT-116 (wt-p53) and HCT-116 (null-p53) cells was size-fractionated using an YM-100 column (Millipore, Billerica, MA) and 0.5 μg size-fractionated RNA were used for the ligation of adaptor sequences. The sequences of the adaptors are as follows: 5′-AAAGGAGGAGCTCTAGaua-3′ and 5′-(P)ugCCTATAGTGAGTCGTATTATTT-3′. Uppercase letters denote deoxyribonucleotides and lower case letters denote ribonucleotides. The adaptors were ligated to the size-fractionated RNA with subsequent gel fractionation steps. Following ligation, the samples were converted to cDNA using a primer complementary to the 3′-adaptor (5′-TAATACGACTCACTATAGGCCA-3′). The cDNA was amplified by PCR using the above-mentioned oligonucleotide as a reverse primer and a forward primer matching the adaptor (5′-AAAGGAGGAGCTCTAGATA-3′). The cDNA was amplified

by PCR and digested with *Xba*I to remove the majority of the 5′ adaptor sequence. The miRNA expression analysis was conducted based on the protocol of Rossetta Genomics (Rehovot, Israel) and Icoria (Research Triangle Park, NC; ref. 22). The array was constructed based on the Sanger Database, containing a total of 247 known miRNAs. cDNA labeled with either Cy3-CTP or Cy5-CTP was generated from HCT-116 (wt-p53) and HCT-116 (null-p53) using the low-input linear amplification kit (Agilent) according to the protocol of the manufacturer. Hybridized microarrays were scanned using the Agilent LP2 DNA Microarray Scanner at 10 μm resolution. Microarray images were visually inspected for defects. The expression of miRNAs was analyzed using Feature Extraction Software (Agilent). The signal of each probe was set as its median intensity. The threshold for reliable probe signals was set at 1,500. Clustering analysis was done using CLUSTER 3.0/TreeView software (23).

Real-time quantitative reverse transcription-PCR analysis for mRNA expression. Real-time quantitative reverse transcription-PCR (qRT-PCR) analysis was done using total RNAs isolated from HCT-116 (wt-p53) and HCT-116 (null-p53) cells and RNAs isolated from both cell lines treated with 10 μmol/L 5-FU for 24 hours. Real-time qRT-PCR primers and probes for *p53* and *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) were purchased from Applied Biosystems, Inc. (Foster City, CA). qRT-PCR was done on an ABI 7500HT instrument under the following conditions: 25°C, 10 minutes; 37°C, 2 hours for reverse transcription; and 95°C, 10 minutes; 95°C, 15 seconds; 60°C, 1 minute for PCR. The reaction was done up to 40 cycles (n = 3). The gene expression ΔC_T value of p53 from each sample was calculated by normalizing with internal house keeping gene *GAPDH* and relative quantitation values were plotted.

Real-time qRT-PCR analysis for miRNA expression. Real-time qRT-PCR analysis was done using total RNAs isolated from HCT-116 (wt-p53) and HCT-116 (null-p53) cells and RNAs isolated from both cell lines treated with 10 μmol/L 5-FU for 24 hours. The miRNA sequence-specific RT-PCR primers for hsa-miR-30a-5p, hsa-miR-181b, hsa-let-7g, hsa-miR-26a, hsa-let-7b, has-miR-15b, has-miR-27a, has-miR-200c, has-miR-191, has-miR-30c, and endogenous control 5S rRNA were purchased from Ambion (Austin, TX). Real-time qRT-PCR analysis was done on an ABI 7500HT instrument using *mir*Vana qRT-PCR miRNA Detection kit (Ambion) under the following conditions: 37°C, 30 minutes; 95°C, 10 minutes of reverse transcription; 95°C, 3 minutes; 95°C, 15 seconds; 60°C, 35 seconds. The reaction was done up to 40 cycles (n = 3). The gene expression ΔC_T values of miRNAs from each sample were calculated by normalizing with internal control 5S rRNA and relative quantitation values were plotted.

Decreasing p53 expression via siRNA knockdown. siRNA molecules were purchased from Dharmacon Research (Lafayette, CO), including p53, positive control (Lamin A/C), and mismatch control. Oligofect-AMINE-mediated transfection of siRNA was carried out in six-well tissue culture plate according to instructions of the manufacturer (Invitrogen). Transfection mixtures containing either 100 or 400 nmol/L siRNA and 8 μL OligofectAMINE in 200 μL Opti-MEM (Invitrogen) were added directly to preincubated cells in 800 μL Opti-MEM. Cells were then incubated for 4 hours and cultured further in McCoy's medium supplemented with 10% fetal bovine serum. Cells were harvested after 48 hours of transfection and total cellular proteins were isolated for Western immunoblot analysis.

Western immunoblot analysis. Western immunoblot analysis was used to characterize the expression of p53 protein after gene knockdown by siRNA and 5-FU treatment in HCT-116 (wt-p53) cells and HCT-116 (null-p53) cells. Equal amounts (15 μg) of protein extracts from each sample were resolved by SDS-PAGE on 12.5% gels by the method of Laemmli (24). Proteins were probed with mouse anti-p53 monoclonal antibody (1:1,000 dilution), α-tubulin (1:3,000 dilution, Santa Cruz Biotechnology, Santa Cruz, CA) followed by incubation with a horseradish peroxidase–conjugated secondary antibody (1:1,000 dilution, Bio-Rad, Hercules, CA). Proteins were visualized with a chemiluminescence detection system using the Super Signal substrate (Pierce, Rockford, IL).

Identification of putative p53-binding site(s) at the miRNA promoters. To identify potential p53-binding sites related to human miRNAs, a set of putative miRNA promoters were extracted by defining 5 kb upstream region of each miRNA precursor. The miRNA genomic coordinates of 326 annotated miRNAs were identified from the miRBase (25). In contrast to protein coding gene, where 1 to 2 kb immediately upstream of the transcription start site are usually used as promoters, instead we chose a 5 kb region upstream of each miRNAs because it is well known that the nuclear transcripts of miRNAs are longer than the known pre-miRNA hairpin precursor that is documented in the databases, and therefore transcription start sites still remain undefined.

Results and Discussion

In this study, a comprehensive analysis was provided for gene expression regulated by wt-p53 at multilevels using human colon cancer cell lines HCT-116 (wt-p53) and HCT-116 (null-p53). This includes steady-state total mRNAs, actively translated mRNAs, and small noncoding miRNAs. The global regulatory network regulated by wt-p53 was revealed, which included transcription, posttranscription, and translation. HCT-116 (wt-p53) and HCT-116 (null-p53) cell lines provide a well-controlled

Table 1. Expression analysis of miRNAs using MIRCHIP2 array in HCT-116 (wt-p53) and HCT-116 (null-p53) cells

Up-regulated miRNA	Fold-change	Down-regulated miRNA	Fold change
Hsa-miR-30a-5p	+32.56	Hsa-miR-15b	−126.34
Hsa-miR-181b	+11.87	Hsa-miR-27a	−60.76
Hsa-miR-372	+6.57	Hsa-miR-200c	−58.30
Hsa-let-7g	+4.93	Hsa-miR-191	−56.39
Hsa-miR-26a	+2.27	Hsa-miR-30c	−50.12
Hsa-let-7b	+2.26	Hsa-miR-25	−49.24
Hsa-miR-296	+2.26	Hsa-miR-107	−48.55
Hsa-miR-30a-3p	+2.00	Hsa-miR-339	−46.52
Hsa-miR-21	+1.89	Hsa-miR-125a	−39.97
Hsa-miR-132	+1.58	Hsa-miR-27b	−32.84
Hsa-miR-181a	+1.34	Hsa-miR-23a	−28.88
Hsa-miR-320	+1.05	Hsa-miR-10a	−3.51

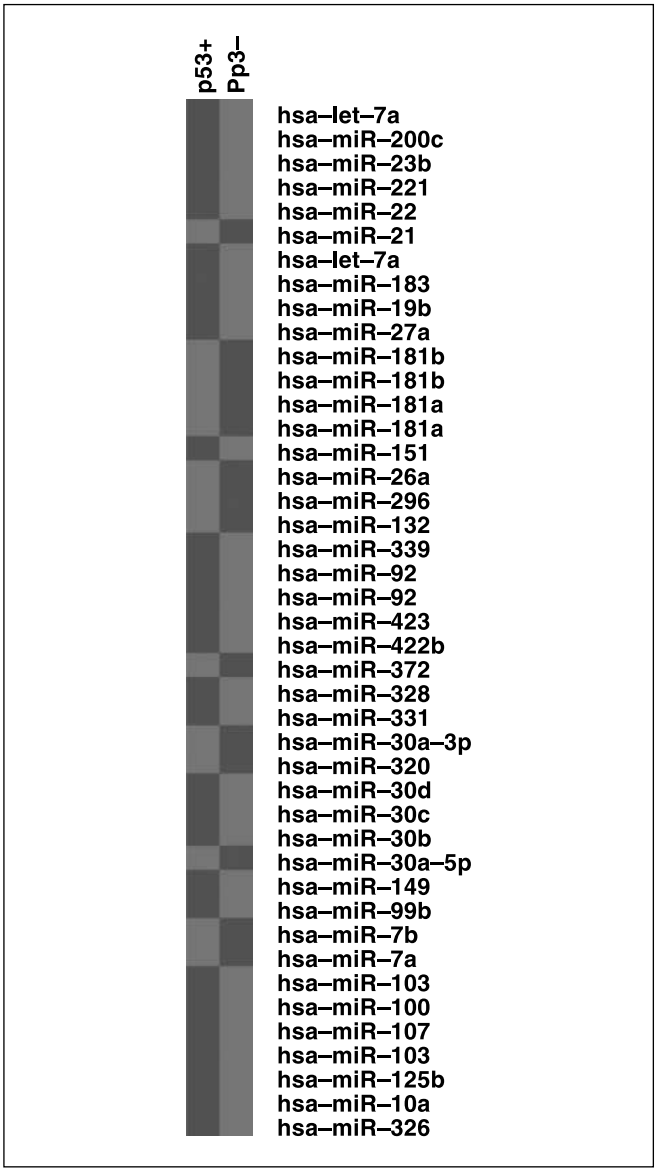


Fig. 1. Hierarchical clustering analysis of miRNA expression (*light gray,* overexpressed genes; *dark gray* underexpressed genes). miRNAs were isolated from actively translated RNA population in HCT-116 (wt-p53) and HCT-116 (null-p53) cells using gel fractionation. The level of miRNAs were quantified with miRNA array analysis. The expression of miRNAs were analyzed using Feature Extraction Software. The signal of each probe was set as its median intensity. The threshold for reliable probe signals was set at 1,500. Clustering analysis was done using CLUSTER 3.0/TreeView software (23).

and ideal *in vitro* model. The *wt-p53* gene in HCT-116 (wt-p53) cells was completely inactivated by targeted deletion using homologous recombination to create HCT-116 (null-p53) cells (15). Although the rate of cell proliferation seemed to be similar in HCT-116 (p53-null) cells compared with HCT-116 (wt-p53) cells in culture, knocking out wt-p53 has already provided HCT-116 (null-p53) cells with certain potential growth advantages under stress conditions. In this report, we discovered that the expression of a number of miRNAs and mRNAs have been affected by the deletion of wt-p53 in HCT-116 (null-p53) cells. We believe some of these altered miRNA and mRNA expression will provide surviving advantage to the HCT-116 (null-p53) cells after genotoxic stress.

Due to the function of miRNAs as translational regulators, we have reasoned that the active population of miRNAs must be localized in the polysomes. The miRNAs were isolated from actively translated RNA population using gel fractionation and the level of miRNA expression was quantitated with miRNA array analysis. We found that 11 miRNAs were up-regulated by wt-p53 and nearly 43 miRNAs were down-regulated by wt-p53 (Table 1). Hierarchical clustering analysis of miRNA expression is shown in Fig. 1. The large number of down-regulated miRNAs is intriguing because it has been predicted that some miRNAs might function as oncogenes due to their suppressive activity (26). We speculate that, as an RNA-binding protein, p53 might affect the recruitment of certain miRNA molecules to the actively translated mRNAs complex. RNA-binding protein tends to interact with a conserved stem-loop secondary structure rather than conserved sequence (27). This is consistent with the fact that most of the miRNAs contain conserved stem-loop structure. On the other hand, p53 acts as a transcription factor to up-regulate certain miRNAs and many downstream cellular mRNAs, including cyclin-dependent kinase inhibitor *p21* gene expression during genotoxic stress. Based on various miRNA target prediction algorithms, it is predicted that roughly 30% of all genes are regulated by miRNAs (28). The prediction points out the potential functional significance of wt-p53-mediated noncoding miRNAs. Several up-regulated miRNAs, such as hsa-miR-181b and hsa-miR-132, have been shown to alter the process of cell proliferation (29). Hsa-miR-21 was shown in a recent report to play a role in regulating apoptosis in human glioblastoma cells (13). The down-regulated miRNAs by antisense against hsa-miR-191 caused increased cell proliferation in HeLa cells, which contain a p53 deletion. In contrast, down-regulating hsa-miR-191 in A549 human lung cancer cells decreased cell proliferation (29). We analyzed the status of p53 in A549 cell lines and the results indicated that A549 cells contain *wt-p53* gene. It seems likely, therefore, that there might be a

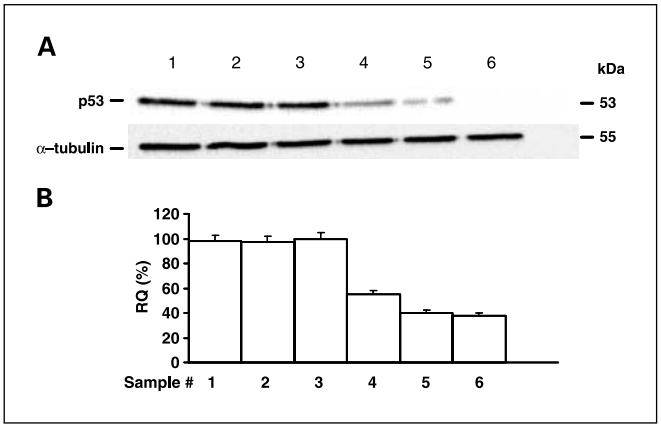


Fig. 2. *A,* effect of p53 siRNA knockdown on the expression of wt-p53 via Western immunoblot analysis. HCT-116 (wt-p53) cells were transfected with siRNA mismatch control and wt-p53-specific siRNA and the expression levels of wt-p53 were quantified via Western immunoblot analysis [*lane 1,* control; *lane 2,* mismatch control; *lane 3,* Lamin A/C positive control; *lane 4,* 100 nmol/L siRNA; *lane 5,* 400 nmol/L siRNA; *lane 6,* HCT-116 (null-p53)]. *B,* real-time qRT-PCR analysis of has-miR-26a expression [$n = 3$; *lane 1,* control; *lane 2,* mismatch control; *lane 3,* Lamin A/C positive control; *lane 4,* 100 nmol/L siRNA; *lane 5,* 400 nmol/L siRNA; *lane 6,* HCT-116 (null-p53)]. 5S rRNA was used as internal standard for expression normalization.

Table 2. Prediction of putative p53-binding sites of miRNA promoters

miRNA	Site position (upstream miRNA)	Gap	Sequence	Site score
Shorter gap sites				
Hsa-let-7b	828	0	AGCCATGTCT...CTTCTTGTCT	87.56
Hsa-mir-26a-1	3,108	2	CAGCAAGACT...GGGCAAGAGC	86.96
Hsa-mir-26a-2	3,024	0	GCCCTTGCCC...CTGCTTGTCT	86.30
Hsa-mir-372	3,661	3	CGCCATGTTG...AGGCTAGTCT	84.81
Hsa-let-7b	3,628	1	TCGCATGCCT...TGTCTTGCTG	83.73
Hsa-mir-181b-1	1,420	0	AGCCAAGCTT...TGGCATGACT	82.44
Hsa-mir-200c	2,183	2	AGACAAGGAG...GAGCAAGGGT	81.59
Hsa-mir-26a-2	174	3	CAGCATGTTG...AGTCAAGTTC	80.06
Longer gap sites				
Hsa-mir-200c	3,807	11	ATACAAGCCG...AGGCAAGTCC	89.80
Hsa-mir-181b-1	1,470	7	AAACATGTCC...CAACTTGCCT	89.00
Hsa-mir-181b-1	4,594	6	GAACTAGCCC...GGCCATGTTT	88.70
Hsa-mir-26a-2	4,840	13	AAGCAAGCAC...GAGCAAGACT	87.77
Hsa-let-7g	1,620	11	AGGCTTGCCT...CAGCAAGCGC	86.69
Hsa-mir-26a-2	4,066	12	CAGCTTGCTT...TGCCATGCCC	85.20
Hsa-mir-26a-2	4,070	8	TTGCTTGCCC...TGCCATGCCC	84.06
Hsa-mir-26a-1	4,496	6	ACGCAAGTCC...TCCCATGTCC	83.60
Hsa-mir-26a-1	4,496	12	ACGCAAGTCC...GTCCTTGCTT	83.29
Hsa-mir-26a-2	4,846	7	GCACAAGATC...GAGCAAGACT	83.12
Hsa-mir-27a	2,060	11	CCTCATGCCT...GAGCTTGGTT	81.88
Hsa-let-7b	2,624	10	GGGCATGGGG...TAGCATGCCG	80.81
Hsa-mir-26a-2	4,863	12	GAGCAAGACT...TGTCTAGTCT	80.66
Hsa-mir-26a-1	1,437	9	TGCCTTGTTC...GGGCATGCAG	80.52
Hsa-mir-26a-1	4,512	4	TCCCATGTCC...TTTCTTGCTT	80.33
Hsa-mir-181b-2	1,330	6	AAACATGAAT...TGACATGCTG	80.28

connection between the function of hsa-miR-191 and the status of p53. The results may help us to further explain the complex biology and function of miRNAs. It shows that in this case, at least the status of tumor suppressor gene function has to be taken into consideration, not just the expression levels of miRNAs.

To confirm that the expression changes of miRNAs were specific due to the status of the p53, the expression of wt-p53 was reduced in HCT-116 (wt-p53) cells by treatment with p53-specific siRNA. The level of the wt-p53 protein was decreased by 80% after treatment with both 100 and 400 nmol/L siRNA in HCT-116 (wt-p53) cells using Western immunoblot analysis (Fig. 2A). The decrease in p53 expression is sequence specific compared with the siRNA controls (*lanes 1* to *3*, and p53-specific siRNA treatment in *lanes 4* and *5* in Fig. 2A). The expression of selected miRNAs were compared in control HCT-116 (wt-p53) cells and cells treated with p53 siRNA. As an example, decreasing p53 expression via siRNA treatment significantly decreased hsa-miR-26a expression quantified via miRNA-specific qRT-PCR analysis (Fig. 2B). This result confirmed the miRNA array results that difference in hsa-miR-26a expression was directly related to the status of wt-p53 expression.

To investigate the potential function of p53 as a transcription factor of some of these miRNAs, the potential p53-binding sites of miRNA promoters were analyzed using bioinformatics approach. The p53-binding site is a dimer, comprising of two monomers, each is 10 nucleotides long, with a variable spacer that can range between 0 to 13 nucleotides (30). The consensus sequence of the monomer is RRRCWWGYYY (R = G or A, W = T or A, Y = C or T; ref. 31), but there are well-documented sites that deviate from this consensus, such as the p53-binding site in the MDM2 promoter (32). Our program uses the following scheme to scan a given sequence for p53-binding sites: the first step uses TFBS (33) to scan the given sequence for a match to the monomer Position-Specific Scoring Matrix, taken from TRANSFAC (Position-Specific Scoring Matrix accession no. M00761; ref. 34), above a given score S. Then, our program searches for dimer sites (i.e., for pairs of monomer sites with a gap of maximum length G). In addition, the program may screen out dimer sites that deviate from the core consensus of the p53 site. It can disqualify sites on any deviation from the consensus, or allow a single deviation, according to the user's choice.

First, we searched for p53 sites related to the 10 candidate miRNAs: hsa-miR-30a, hsa-miR-181b, hsa-let-7g, hsa-let-7b, hsa-miR26a, hsa-miR-15b, hsa-miR-27a, hsa-miR-200c, hsa-miR-25, and hsa-miR-372 (Table 1). In fact, two of the candidates, hsa-miR-30a and hsa-miR-181b, are each transcribed from two distinct genomic loci, and thus our list of candidate contained 12 promoters. We first used a cutoff score $S = 80$, allowed only short gaps of less than four nucleotides, and demanded that both monomers in each dimer will perfectly match the core consensus. Using these variables, we identified p53-binding sites in 6 of 12 promoters (hsa-miR-181b-1, hsa-let-7b, hsa-miR26a-1, hsa-miR26a-2, hsa-miR-200c, and hsa-miR-372), which correspond to 5 of 10

candidate miRNAs. We relaxed the initial variables and checked for sites in more candidate promoters. Hsa-miR-25 had a perfect consensus site in a score of 78, and another site that deviates from the consensus with a score of 84. In addition, hsa-miR-30a also has a consensus site with a deviation. When we relaxed the gap variable and searched for sites with gap up to 13 nucleotides, two more candidates, hsa-let-7g and hsa-miR-27a, were revealed to contain a perfect consensus site. The results are summarized in Table 2. Overall, we found putative p53 sites for 10 of the 12 candidate promoters, which correspond to 9 of 10 candidate miRNAs that we have checked.

The entire set of 326 miRNA putative promoters were screened for p53 potential binding sites using the first variable configuration (i.e., $S = 80$, $G = 3$) and perfect core-consensus match. The search resulted in 187 sites in the promoters of 130 unique miRNAs. To assess the significance of this result, we repeated the same search on 1,000 sets of 326 reshuffled miRNA promoters. Out of 1,000 reshuffled sets, a mean of 47 and maximum of 69 unique promoters contain at least one p53 site, as opposed to 130 in real promoters. This clearly indicates that relative to randomized versions of the miRNA promoters, the real promoters contain a very high number of p53-binding sites ($P < 0.001$). In summary, the search for p53 sites in the putative promoters of the set of candidate miRNAs resulted in 9 of 10 candidates (or 10 of 12 distinct promoters) containing at least one potential p53-binding site. When looking at the entire set of

Table 3. Expression analysis of steady-state total mRNA transcripts in HCT-116 (wt-p53) and HCT-116 (null-p53) cells

Genbank accession no.	Gene ID	Fold change	Biological function
Increased genes			
NM_000474	*TWIST*	+34.30	Cell differentiation; chromosome organization and biogenesis
NM_001225	*CASP4*	+20.22	Apoptosis; proteolysis and peptidolysis
NM_012427	*KLK5*	+11.13	Epidermis development; proteolysis and peptidolysis
NM_004172	*SLC1A3*	+11.07	L-Glutamate transport; dicarboxylic acid transport
NM_005930	*MGEA6*	+10.92	RNA processing
D16350	*SAH*	+10.07	Metabolism; regulation of blood pressure
NM_004864	*PLAB*	+9.75	Cell-cell signaling; signal transduction
NM_025048	*FLJ22684*	+9.46	Neuropeptide signaling pathway
AB029015	*PLCL2*	+9.11	Intracellular signaling cascade; lipid metabolism
NM_016135	*TEL2*	+9.00	Organogenesis; regulation of transcription
NM_006017	*PROML1*	+8.88	Visual perception
NM_002923	*RGS2*	+8.54	Cell cycle; G-protein signaling pathway; signal transduction
M23419	*EIF5A;TNNI3*	+7.09	Protein biosynthesis; translational initiation
NM_003633	*ENC1*	+6.99	Development; neurogenesis
AK023349	*Nup43*	+4.97	Intracellular protein transport
NM_005596	*NFIB*	+4.89	DNA replication; regulation of transcription, DNA dependent
NM_012198	*GCA*	+3.96	Membrane fusion
NM_001654	*ARAF1*	+3.92	Cell growth and/or maintenance; intracellular signaling cascade
NM_001387	*DPYSL3*	+3.88	Neurogenesis; signal transduction
NM_000043	*TNFRSF6*	+3.80	Apoptosis; immune response; protein assembly; signal transduction
NM_000389	*CDKN1A*	+3.55	Cell cycle; apoptosis; cell proliferation; regulation of CDK activity
BC007613	*CRMP1*	+3.14	Neurogenesis; nucleobase, nucleoside, nucleotide and nucleic acid metabolism
AK057343	*ZNF131*	+3.11	Regulation of transcription, DNA-dependent
NM_052966	*C1orf24*	+3.08	Protein folding
Decreased genes			
NM_014178	*HSPC156*	−21.83	Vesicle-mediated transport
NM_000582	*SPP1*	−19.53	T-helper 1 type immune response; antiapoptosis; cell-cell signaling
NM_001147	*ANGPT2*	−13.04	Angiogenesis; cell growth and/or maintenance; signal transduction
NM_003121	*SPIB*	−11.76	Regulation of transcription from Pol II promoter
BC010398	*PMPCB*	−7.41	Proteolysis and peptidolysis
NM_000310	*PPT1*	−5.26	Neurogenesis; protein modification; visual perception
NM_000465	*BARD1*	−4.59	Protein ubiquitination
NM_016611	*KCNK4*	−4.29	Ion transport; potassium ion transport
NM_001901	*CTGF*	−3.50	DNA metabolism; cell adhesion; cell growth; cell motility
NM_003925	*MBD4*	−3.34	Base-excision repair
NM_005627	*SGK*	−3.09	Apoptosis; phosphorylation; response to stress; sodium ion transport
NM_005834	*TIMM17B*	−3.06	Protein-mitochondrial targeting
NM_013961	*NRG1*	−3.06	Cell differentiation; embryonic development; neurogenesis

miRNAs in the database, we get clear indications that these results are not random and contain a high rate of true positives (estimates at ~64%, when looking at a 1,000 reshuffled sets of promoters as an indication to the rate of false positives). It is, therefore, possible that additional miRNAs may be regulated by p53.

The gene expression of steady-state total mRNA transcripts from both HCT-116 (wt-p53) and HCT-116 (null-p53) cells was

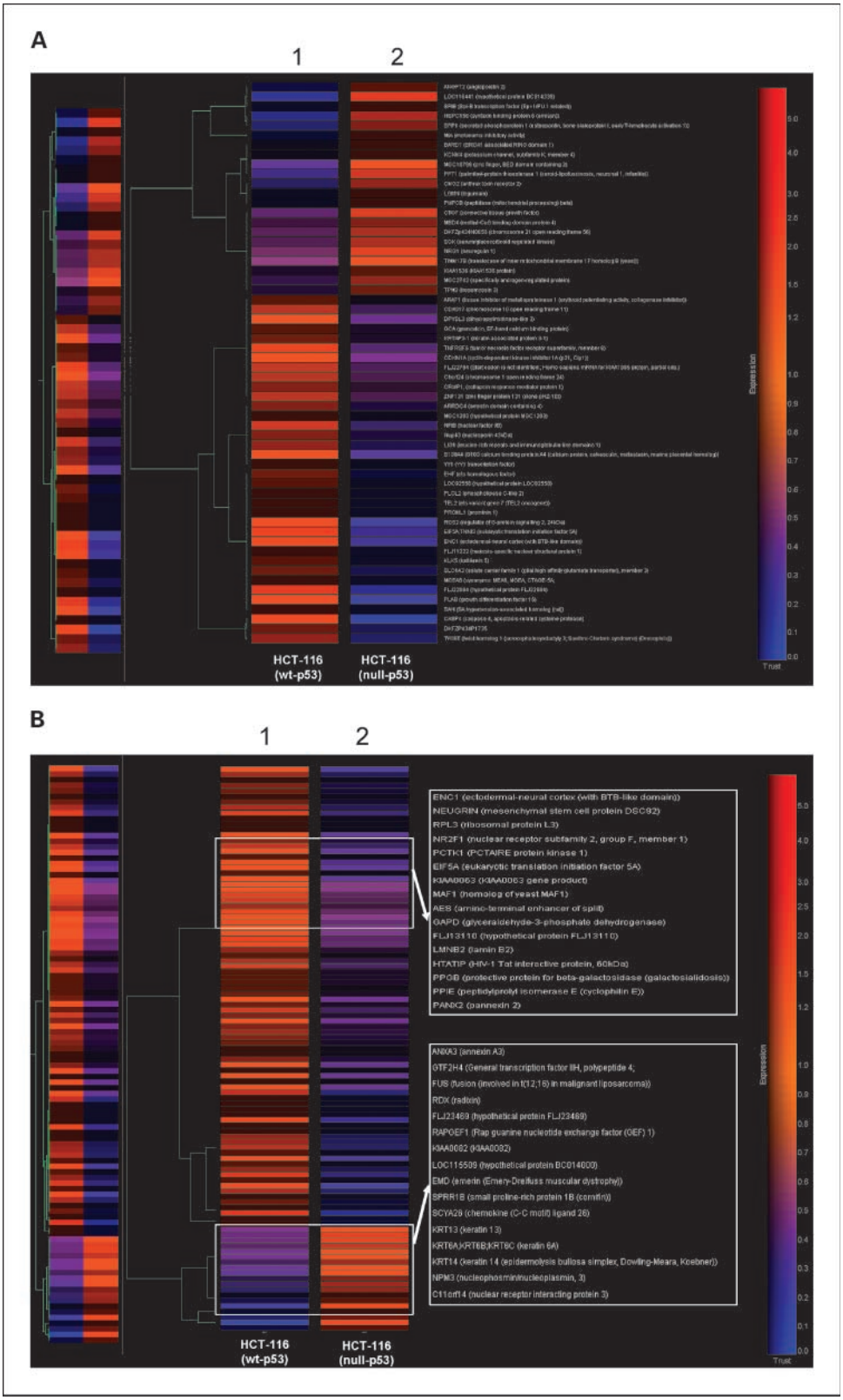


Fig. 3. *A,* hierarchical clustering analysis of steady-state total mRNAs expression profile between HCT-116 (wt-p53; *lane 1*) and HCT-116 (null-p53; *lane 2*) cells via microarray expression analysis (*red,* overexpressed genes; *blue,* underexpressed genes). *B,* hierarchical clustering analysis of actively translated mRNAs expression profile between HCT-116 (wt-p53; *lane 1*) and HCT-116 (null-p53; *lane 2*) cells (*red,* overexpressed genes; *blue,* underexpressed genes).

Table 4. Expression analysis of actively translated mRNA transcripts in HCT-116 (wt-p53) and HCT-116 (null-p53) cells

Genbank accession no.	Gene ID	Fold change	Biological function
Increased genes			
NM_016292	*TRAP1*	+19.01	Protein folding
NM_014474	*ASML3B*	+13.51	Carbohydrate metabolism
NM_001654	*ARAF1*	+12.79	Cell growth and/or maintenance; intracellular signaling cascade
NM_000178	*GSS*	+11.20	Amino acid metabolism; glutathione biosynthesis; neurogenesis
NM_001225	*CASP4*	+8.83	Apoptosis; proteolysis and peptidolysis
NM_001402	*EEF1A1*	+8.44	Protein biosynthesis; regulation of translation
NM_000918	*P4HB*	+7.86	Electron transport
NM_004335	*BST2*	+7.28	Cell proliferation; cell-cell signaling; development; immune response
NM_006743	*RBM3*	+6.85	RNA processing
NM_004046	*ATP5A1*	+6.83	ATP synthesis – coupled proton transport
NM_000546	*TP53*	+6.58	Apoptosis; DNA recombination; DNA repair; cell cycle; transcription
NM_004317	*ASNA1*	+6.54	Anion transport; response to arsenate
NM_004127	*GPS1*	+6.19	c-Jun-NH_2-kinase cascade; cell cycle; inactivation of mitogen-activated protein kinase
NM_000291	*PGK1*	+6.13	Glycolysis
NM_003752	*EIF3S8*	+6.12	Protein biosynthesis; regulation of translational initiation
NM_005500	*SAE1*	+5.85	Protein ubiquitination
NM_000107	*DDB2*	+5.70	Nucleotide-excision repair
AK024835	*CNN2*	+5.68	Cytoskeleton organization; smooth muscle contraction
NM_003915	*CPNE1*	+5.62	Lipid metabolism; vesicle-mediated transport
NM_006374	*STK25*	+5.53	Phosphorylation; response to oxidative stress; signal transduction
NM_017916	*FLJ20643*	+5.44	Metabolism
NM_006400	*DCTN2*	+5.35	Cell proliferation; microtubule-based process; mitosis
NM_000474	*TWIST*	+5.19	Cell differentiation; chromosome organization and biogenesis
NM_005030	*PLK*	+5.09	Mitosis; protein amino acid phosphorylation; regulation of cell cycle
BC033103	*INPP5E*	+5.04	Proteolysis and peptidolysis
NM_032272	*MAF1*	+4.91	Regulation of transcription, DNA-dependent
NM_001970	*EIF5A*	+4.85	Protein biosynthesis; translational initiation
NM_006201	*PCTK1*	+4.73	Protein amino acid phosphorylation; regulation of cell cycle
NM_005654	*NR2F1*	+4.66	Regulation of transcription, DNA-dependent; signal transduction
NM_003633	*ENC1*	+4.62	Development; neurogenesis
NM_016016	*CGI-69*	+4.60	Transport
NM_000967	*RPL3*	+4.58	Protein biosynthesis
NM_016645	*NEUGRIN*	+4.56	Neuron differentiation
NM_018658	*KCNJ16*	+4.38	Ion transport; potassium ion transport
NM_004864	*PLAB*	+4.33	Cell-cell signaling; signal transduction
NM_003624	*RANBP3*	+4.29	Small GTPase-mediated signal transduction
NM_004559	*NSEP1*	+4.28	Regulation of transcription; response to pest/pathogen/parasite

analyzed and genes with known functions are listed in Table 3 and hierarchical clustering analysis is shown in Fig. 3A. The list contains many genes involved in cell cycle control (TWIST, CASP4, and CDKN1A) and altogether 63 genes were affected by the deletion of wt-p53. It is interesting to note that the expression of SPIB, a regulator of transcription from Pol II promoter, is decreased by 11-fold in HCT-116 (wt-p53) cells. It has been reported that transcription of miRNAs are mediated by RNA polymerase II (35), which could help to explain another potential regulatory mechanism of miRNAs with decreased expressions listed in Table 1.

As an RNA-binding protein, p53 regulates gene expression at the posttranscriptional level. miRNAs also regulate gene expression at posttranscriptional level. The changes in the rate of mRNA translation will not be captured by just quantifying steady-state total mRNA levels. Therefore, it is critical to analyze gene expression using actively translated mRNA transcripts. The gene expression profiles of actively translated mRNA transcripts from both HCT-116 (wt-p53) and HCT-116 (null-p53) cells were analyzed, and genes with known functions are listed in Table 4. Hierarchical clustering analysis is shown in Fig. 3B. The results indicated that 107 genes were affected at the level of posttranscriptional control, many of which are related to RNA processing (RBM3) and protein synthesis (EEF1A1, EIF3S8, EIF5A, and EIF4A; Table 4). These changes may be mediated by p53 at the posttranscriptional level via direct or indirect manner and some of the genes that were indirectly regulated posttranscriptionally

Table 4. Expression analysis of actively translated mRNA transcripts in HCT-116 (wt-p53) and HCT-116 (null-p53) cells (Cont'd)

Genbank accession no.	Gene ID	Fold change	Biological function
NM_012427	*KLK5*	+4.14	Epidermis development; proteolysis and peptidolysis
NM_001085	*SERPINA3*	+4.12	Inflammatory response; regulation of lipid metabolism
NM_021734	*SLC25A19*	+4.12	Deoxynucleotide transport
NM_000175	*GPI*	+4.10	Carbohydrate metabolism; gluconeogenesis; glycolysis; hemostasis
NM_001908	*CTSB*	+4.10	Proteolysis and peptidolysis
NM_002923	*RGS2*	+4.01	Cell cycle; G-protein signaling pathway; signal transduction
NM_000687	*AHCY*	+3.95	One-carbon compound metabolism
NM_000190	*HMBS*	+3.94	Heme biosynthesis
NM_013388	*PREB*	+3.93	Regulation of transcription, DNA dependent
NM_005410	*SEPP1*	+3.92	Response to oxidative stress
NM_015679	*CLONE24922*	+3.87	RNA processing
NM_021008	*DEAF1*	+3.82	Regulation of transcription, DNA dependent
AK000822	*DKFZP564M182*	+3.75	Protein biosynthesis
NM_000182	*HADHA*	+3.74	Fatty acid metabolism
NM_002611	*PDK2*	+3.70	Glucose metabolism; signal transduction
NM_005802	*TP53BPL*	+3.68	Protein ubiquitination
AF016266	*TNFRSF10B*	+3.59	Caspase activation; electron transport; induction of apoptosis
NM_001536	*HRMT1L2*	+3.58	Signal transduction; defense response; methylation
NM_133455	*LOC129080*	+3.52	Phosphate transport
NM_001640	*APEH*	+3.51	Proteolysis and peptidolysis
NM_001416	*EIF4A1*	+3.50	Protein biosynthesis
NM_006388	*HTATIP*	+3.33	Regulation of transcription, DNA-dependent; transcription
NM_006112	*PPIE*	+3.29	Protein folding
NM_000308	*PPGB*	+3.28	Intracellular protein transport; proteolysis and peptidolysis
NM_002046	*GAPD*	+3.24	Glucose metabolism; glycolysis
NM_001130	*AES*	+3.24	Wnt receptor signaling pathway; development; transcription
Decreased genes			
NM_014178	*HSPC156*	−27.25	Vesicle-mediated transport
NM_006533	*MIA*	−15.34	Cell proliferation
NM_006993	*NPM3*	−7.58	Protein folding
NM_005554	*KRT6A*	−6.62	Ectoderm development
NM_002274	*KRT13*	−5.75	Epidermis development
NM_003125	*SPRR1B*	−4.81	Epidermis development
NM_000117	*EMD*	−4.76	Muscle contraction; muscle development
NM_006072	*SCYA26*	−4.69	Cell-cell signaling; immune response; signal transduction
BC014000	*LOC115509*	−4.18	Regulation of transcription, DNA dependent
NM_024710	*FLJ23469*	−3.88	Metabolism
NM_002906	*RDX*	−3.47	Cytoskeletal anchoring

may be mediated via miRNAs. We attempted to match the potential mRNA targets with several miRNAs using predictive software (miRNada; refs. 28, 36), and the results are shown in Fig. 4. These results were based on the improved prediction rules by reducing the number of G:U wobbles and increase the high match scale factor from 2 to 4 at position 2-8 from the 5′ end of miRNA (rather than position 1-11). Interestingly, the expression levels of translation initiation factor 4A and 5A were altered by the expression of p53. This change may be mediated by miRNAs such as hsa-miR-15b and hsa-miR-125a, respectively, based on target prediction analysis. It is possible that miRNAs are joining force with wt-p53 to help regulating gene expression at multiple levels.

To further confirm the functional significance of the p53 in regulating translation, both HCT-116 (wt-p53) and HCT-116 (null-p53) cells were treated with 10 μmol/L 5-FU for 24 hours. 5-FU is one of the main anticancer compounds used in treating colorectal cancer. The regulation of p53 was known to be controlled at the posttranscriptional level (37). To validate our approach, the expression level of p53 was analyzed by Western immunoblot and real-time qRT-PCR analysis. The wt-p53 protein level was induced after 5-FU treatment in HCT-116 (wt-p53) cells (Fig. 5A). However, the level of wt-p53 mRNA was not changed by 5-FU treatment (Fig. 5B). These results, taken together, suggest that the up-regulation of wt-p53 after 5-FU exposure is indeed due to posttranscriptional regulation. These data clearly point out the importance of

```
3'_ACAUUUGGUACUA-C-ACGACGAU_5':hsa-miR-15b
___||.|.:|.:||.|.|.||:||||:___
5'_TGGATGCAGTGCTCGCTGTTGCTG_3': eIF-4A

3'_gugUCCAAUUUCCCAGAGUCCcu_5':hsa-miR-125a
___...|.||.|.||.|.|||||||..___
5'_gcgATGTCATAGCGGCTCAGGat_3': Heat Shock Protein 5A

3'_cUUUCUCU-GGCCAAGUGACACu_5':hsa-miR-128b
___.:||||||.||.|.||||.|||.___
5'_aGAAGAGATCCTGATCACGGTGc_3': eIF-5A

3'_gUGUUUAAGCC---UAGAUGUCCCAu_5':hsa-miR-10a
___.:|.|:.|.||...:|.||:|||||.___
5'_aGCCAGGTGGGCCTGTATATAGGGTc_3': Protein phosphatase 1,
                                  subunit 12C
```

Fig. 4. miRNA target prediction for hsa-miR-15b, hsa-miR-125a, hsa-miR-128b, and hsa-miR-10a. Potential mRNA targets with several miRNAs using predictive software (miRNada; refs. 28, 36).

analyzing gene expression using actively translated mRNAs in addition to the steady-state total mRNAs. Several miRNAs was also up-regulated in response to wt-p53 induction after 5-FU treatment based on real-time qRT-PCR analysis. The expression analysis of hsa-miR-26a in response to 5-FU treatment is shown in Fig. 5C. The expression level of hsa-miR-26a was increased by nearly 2-fold in HCT-116 (wt-p53) cells in response to the increasing expression of wt-p53 after 5-FU exposure. In contrast, there was only a slight increase in HCT-116 (null-p53) cells after 5-FU treatment. These results further support the functional significance of wt-p53 on miRNA expression.

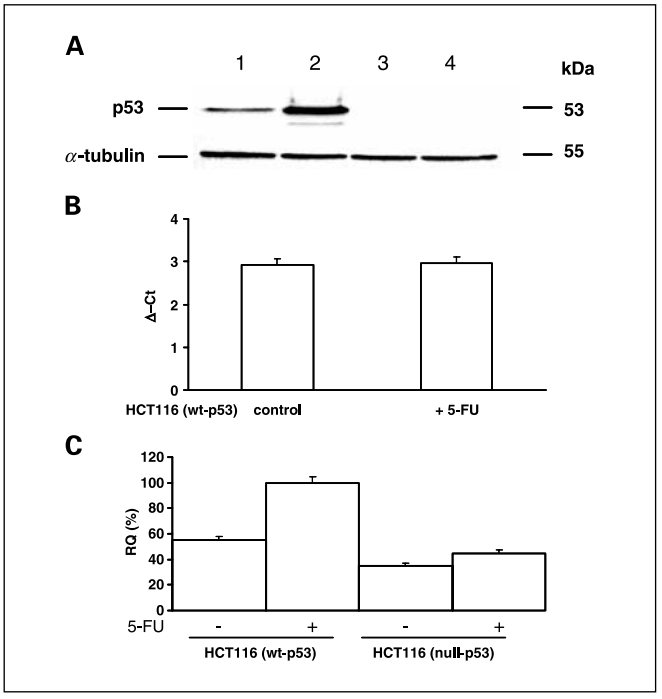


Fig. 5. *A,* effect of 5-FU treatment on the expression of p53 via Western immunoblot analysis [*lane 1,* control HCT-116 (wt-p53); *lane 2,* HCT-116 (wt-p53) cells treated with 10 μmol/L 5-FU for 24 hours; *lane 3,* control HCT-116 (null-p53); *lane 4,* HCT-116 (null-p53) cells treated with 10 μmol/L 5-FU for 24 hours]. *B,* expression of wt-p53 mRNA from HCT-116 (wt-p53) control cells and cells treated with 10 μmol/L 5-FU for 24 hours analyzed by real-time qRT-PCR analysis ($n = 3$). The expression of housekeeping gene *GAPDH* was used as internal control. *C,* effect of 5-FU treatment on hsa-miR-26a expression analyzed by real-time miRNA qRT-PCR ($n = 3$) and 5S rRNA was used as internal standard. The ratio of target miRNA and 5S rRNA was used to calculate the relative expression.

Based on our results, we provide a flow diagram to illustrate the new aspects of the regulatory function of wt-p53 (Fig. 6) to better understand the complexity of the regulatory network mediated by wt-p53. Wt-p53 not only regulates posttranscriptional and translational events via its RNA-binding function, but also acts as a typical transcription factor to regulate a number of cellular mRNAs at the transcriptional level. The new aspect of this regulatory network is that wt-p53 also regulates a number of noncoding miRNAs at the transcriptional level. Therefore, it is very likely that wt-p53 also regulates certain cellular mRNA translation through its mediated miRNAs. Wt-p53 also enhances apoptosis via directly accumulating to mitochondria (4, 5).

In conclusion, we describe here a comprehensive gene expression analysis to provide evidence that wt-p53 regulates gene expression at multiple levels due to its diverse functions. Wt-p53 not only regulates gene expression as a transcription factor to induce mRNA expression, but also influences miRNA expression by direct or indirect manner in this colon cancer cell line model. Wt-p53 also affects gene expression at posttranscriptional levels either through miRNAs or its RNA-binding capability. Some of the miRNAs have been shown to play roles in cell proliferation and apoptosis (29). We have also identified a number of wt-p53-regulated mRNAs at both transcriptional and posttranscriptional levels and some of these genes are candidate targets for miRNAs. The understanding of the complicated molecular networks regulated by wt-p53 is crucial in further elucidation of gene regulation.

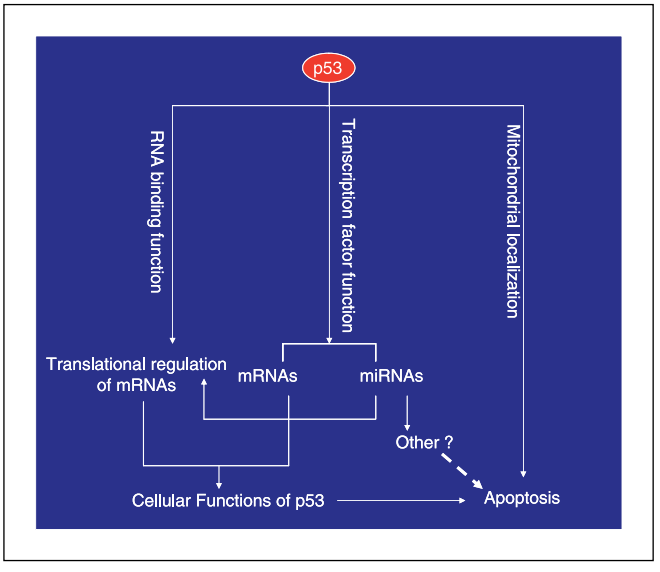


Fig. 6. Schematic illustration of p53 regulatory pathways. Wt-p53 not only acts as a typical transcription factor to regulate a number of cellular mRNAs at the transcriptional level, but also regulates gene expression at posttranscriptional and translational events via its RNA-binding function. In addition, wt-p53 regulates a number of noncoding miRNAs at the transcriptional level thereby influences certain cellular mRNAs translation through its mediated miRNAs. Wt-p53 also influences apoptosis pathway via accumulation to the mitochondria.

References

1. Waldman T, Kinzler KW, Vogelstein B. p21 is necessary for the p53-mediated G_1 arrest in human cancer cells. Cancer Res 1995;55:5187–90.
2. Fu L, Minden MD, Benchimol S. Translational regulation of human p53 gene expression. EMBO J 1996;15: 4392–401.
3. Miller SJ, Suthiphongchai T, Zambetti GP, Ewen ME. p53 binds selectively to the 5′ untranslated region of cdk4, an RNA element necessary and sufficient for transforming growth factor β- and p53-mediated translational inhibition of cdk4. Mol Cell Biol 2000; 20:8420–31.
4. Donahue RJ, Razmara M, Hoek JB, Knudsen TB. Direct influence of the p53 tumor suppressor on mitochondrial biogenesis and function. FASEB J 2001;15: 635–44.
5. Marchenko ND, Zaika A, Moll UM. Death signal-induced localization of p53 protein to mitochondria. A potential role in apoptotic signaling. J Biol Chem 2000;275:16202–12.
6. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75: 843–54.
7. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 1993;75:855–62.
8. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC. Expression of *Arabidopsis* miRNA genes. Plant Physiol 2005;138:2145–54.
9. Sheikh MS, Fornace AJ, Jr. Regulation of translation initiation following stress. Oncogene 1999;18: 6121–8.
10. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature 2005;435:839–43.
11. Ju J, Pedersen-Lane J, Maley F, Chu E. Regulation of p53 expression by thymidylate synthase. Proc Natl Acad Sci U S A 1999;96:3769–74.
12. Feliers D, Duraisamy S, Barnes JL, Ghosh-Choudhury G, Kasinath BS. Translational regulation of vascular endothelial growth factor expression in renal epithelial cells by angiotensin II. Am J Physiol Renal Physiol 2005;288:F521–9.
13. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res 2005;65:6029–33.
14. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 2002;99:15524–9.
15. Bunz F, Dutriaux A, Lengauer C, et al. Requirement for p53 and p21 to sustain G_2 arrest after DNA damage. Science 1998;282:1497–501.
16. Bunz F, Hwang PM, Torrance C, et al. Disruption of p53 in human cancer cells alters the responses to therapeutic agents. J Clin Invest 1999;104:263–9.
17. Idelman G, Glaser T, Roberts CT, Jr., Werner H. WT1-53 interactions in insulin-like growth factor-I receptor gene regulation. J Biol Chem 2003;278:3474–82.
18. Yu JL, Rak JW, Coomber BL, Hicklin DJ, Kerbel RS. Effect of p53 status on tumor response to antiangiogenic therapy. Science 2002;295:1526–8.
19. Kim J, Krichevsky A, Grad Y, et al. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci U S A 2004; 101:360–5.
20. Ju J, Huang C, Minskoff SA, Mayotte JE, Taillon BE, Simons JF. Simultaneous gene expression analysis of steady-state and actively translated mRNA populations from osteosarcoma MG-63 cells in response to IL-1α via an open expression analysis platform. Nucleic Acids Res 2003;31:5157–66.
21. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 2001;15:188–200.
22. Barad O, Meiri E, Avniel A, et al. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res 2004;14:2486–94.
23. Chiang DY, Brown PO, Eisen MB. Visualizing associations between genome sequences and gene expression data using genome-mean expression profiles. Bioinformatics 2001;17 Suppl 1:S49–55.
24. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680–5.
25. Griffiths-Jones S. The microRNA registry. Nucleic Acids Res 2004;32:D109–11.
26. McManus MT. MicroRNAs and cancer. Semin Cancer Biol 2003;13:253–8.
27. Chu E, Copur SM, Ju J, et al. Thymidylate synthase protein and p53 mRNA form an *in vivo* ribonucleoprotein complex. Mol Cell Biol 1999;19:1582–94.
28. Krek A, Grun D, Poy MN, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37: 495–500.
29. Cheng AM, Byrom MW, Shelton J, Ford LP. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. Nucleic Acids Res 2005;33:1290–7.
30. el-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B. Definition of a consensus binding site for p53. Nat Genet 1992;1:45–9.
31. Wang L, Wu Q, Qiu P, et al. Analyses of p53 target genes in the human genome by bioinformatic and microarray approaches. J Biol Chem 2001;276: 43604–10.
32. Zauberman A, Barak Y, Ragimov N, Levy N, Oren M. Sequence-specific DNA binding by p53: identification of target sites and lack of binding to p53-MDM2 complexes. EMBO J 1993;12:2799–808.
33. Lenhard B, Wasserman WW. TFBS: Computational framework for transcription factor binding site analysis. Bioinformatics 2002;18:1135–6.
34. Matys V, Fricke E, Geffers R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 2003;31:374–8.
35. Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J 2004;23: 4051–60.
36. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human microRNA targets. PLoS Biol 2004; 2:e363.
37. Dony C, Kessel M, Gruss P. Post-transcriptional control of myc and p53 expression during differentiation of the embryonal carcinoma cell line F9. Nature 1985;317:636–9.